Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Ярославский государственный университет
им. П.Г. Демидова

В.Е. Орёл
И.Г. Сенин

Основы психодиагностики

Учебное пособие
Издание второе, переработанное и дополненное

*Рекомендовано Научно-методическим советом
университета для студентов, обучающихся
по специальности Психология*

Ярославль 2007

УДК 159.9.07.(075.8)
ББК Ю 93я73
О 65

Рекомендовано
Редакционно-издательским советом университета
в качестве учебного издания. План 2007 года

Рецензенты:
А.О. Прохоров, доктор психологических наук, профессор, зав. кафедрой общей психологии Казанского государственного университета;
кафедра общей и социальной психологии Ярославского государственного педагогического университета им. К.Д. Ушинского

Орёл, В.Е. Основы психодиагностики: учеб. пособие / В.Е. Орёл,
О 65 И.Г. Сенин; Яросл. гос. ун-т. – 2-е изд., перераб. и доп. – Ярославль : ЯрГУ, 2007. – 140 с.
ISBN 978-5-8397-0558-6

Рассматриваются основные теоретические и прикладные проблемы психодиагностики. Дается анализ различных теоретических подходов к предмету психодиагностики, раскрывается содержание этапов разработки, создания и проверки психологических тестов. При подготовке издания авторы опирались на собственный опыт создания, проверки и использования тестовых методик.

Во втором издании существенно расширен ряд разделов, которые касаются основных методов математической статистики, проверки надежности и валидности теста. Включены новые разделы, касающиеся стандартизации процедуры и условий тестирования и адаптации зарубежных тестов. Пособие дополнено статистическими таблицами.

Предназначено студентам, обучающимся по специальности 030301 Психология (дисциплина «Психодиагностика», блок ОПД), очной формы обучения, может быть также рекомендовано практическим психологам.

Рис. 4. Табл. 2. Библиогр.: 69 назв.

УДК 159.9.07.(075.8)
ББК Ю 93я73

ISBN 978-5-8397-0558-6

Оглавление

Светлой памяти
Алексея Рукавишникова,
соратника и друга,
посвящается эта книга

Введение

Психологическая практика невозможна без применения методов диагностики психических характеристик личности, дающих основание для проведения консультационной и коррекционной работы, различного рода психологических экспертиз, а также научных исследований в различных областях психологии. Поэтому тест как метод исследования личности занимает прочные позиции в методическом арсенале психолога. Это обусловило появление не только большого числа тестовых методик в отечественной психологии, но и привлекло внимание специалистов к вопросам методологии разработки и использования тестов. Конкретным проявлением этого стало появление ряда отечественных публикаций, посвященных проблемам тестирования и ставших своего рода классикой в этой области [3], [7], [21], [35], а также перевод известных зарубежных источников [2], [12].

Вместе с тем, опыт работы многих отечественных психологов, а также наш собственный опыт в области психодиагностики показывает, что не все пользователи психологических тестов имеют адекватное представление об особенностях создания тестов, критериях оценки их эффективности, этических принципах проведения тестирования; у большинства пользователей отсутствуют знания о критериях, позволяющих отличить «хороший» тест от «плохого». Отражением этой ситуации является появление на полках книжных магазинов большого количества популярной литературы, в которой публикуются не только тесты для самопознания, но и методики, работа с которыми возможна только при наличии у пользователя психологической квалификации. Данная ситуация в области практической психодиагностики не

только не способствовала повышению эффективности практической работы, но и создавала предпосылки для дискредитации психологии как науки. Все это послужило одной из причин написания учебного пособия. Другая причина заключалась в повышении требований к выпускникам психологических факультетов в плане наличия у них соответствующих знаний, умений и навыков в области конструирования, проверки и использования тестовых методик, что нашло отражение в принятом Государственном образовательном стандарте.

Авторы поставили перед собой задачу осветить проблемы теории и практики психодиагностики, рассмотреть вопросы ее предмета, проследить основные этапы процедуры создания теста, принятой в традиционной психодиагностике. В подготовке и написании данного учебного пособия мы ориентировались на труды известных в этой области отечественных и зарубежных исследователей: А. Анастази, Л.Ф. Бурлачука, Г. Витцлака, В.Н. Дружинина, П. Клайна, А.Г. Шмелева и других. В публикуемой работе нашел отражение и собственный, более чем десятилетний опыт работы авторов по созданию и адаптации тестовых методик, проводимой в рамках деятельности научно-производственного центра «Психодиагностика».

Подготовка нового, второго издания пособия была вызвана несколькими причинами: 1. Преподавание курса «Психодиагностика» в течение нескольких лет после выхода первого издания вскрыло ряд слабых мест в подготовке будущих психологов, особенно в использовании математико-статистических методов и понимании основных психометрических характеристик. 2. Некоторые разделы, наиболее сложные для понимания, требовали иллюстраций примерами, что облегчало их восприятие и понимание. 3. Первое издание пособия достаточно быстро разошлось среди студентов, создав противоречие между их потребностями в учебной литературе и реальными возможностями.

Структура работы включает введение, 6 глав и заключение. Введение, главы 1, 2, 3, 4, 5 (разделы 6 и 7), глава 6 и заключение написаны доктором психологических наук, профессором В.Е. Орлом, глава 5 (разделы 1, 2, 3, 4, 5) кандидатом психологических наук, доцентом И.Г. Сениным

Нет умных и глупых тестов.
Есть умное и глупое их применение.
Б.М. Теплов

Глава 1. **Предмет и задачи психодиагностики**

1.1. Предмет психодиагностики

Термин «психодиагностика» был впервые предложен Г. Роршахом в 1921 году в одноименной работе. Будучи изначально медицинским термином, он постепенно вышел за ее пределы. Вместе с тем, содержание этой области до сих пор не определено, что находит свое отражение и в ее дефинициях. В отечественной психологии под термином «психодиагностика» понимается довольно широкая область методов и методик, призванных описывать психологические различия между людьми и группами людей [35], [39]. С этих позиций психодиагностика предстает некоей методологической основой, объединяющей все области практического применения различных методов и методик: тестов, проективных методик, психофизиологических методов, методов опроса и анамнеза. Как отмечает Г. Витцлак, психодиагностика является общим теоретико-методологическим базисом для оценки личности в системе самых разнообразных отношений [6].

В зарубежной психологии, прежде всего американской, термин «психодиагностика» отождествляется с термином «психологическое тестирование» [53], [57], [61]. С этих позиций психодиагностика рассматривается как дисциплина, связанная с применением разного рода тестов, направленных на выявление индивидуально-психологических различий между людьми. И хотя целый ряд отечественных специалистов и некоторые зарубежные психологи считают такой подход не совсем верным, а психо-

логическое тестирование они рассматривают как часть психодиагностики, в реальной практике работы психолога именно тесты чаще всего используются для постановки психологического диагноза.

Разделяя данную точку зрения, мы будем рассматривать психодиагностику в узком смысле, ***как отрасль психологической науки, связанную с теоретическими основами и практическими аспектами разработки и использования психологических тестов.***

Выделяют два вида психодиагностики: ***общую психодиагностику и частную психодиагностику.***

Общая психодиагностика, являясь как теоретической, так и прикладной дисциплиной, рассматривает закономерности вынесения надежных и валидных диагностических суждений, с помощью которых мы можем оценить наличие и степень выраженности тех или иных психологических характеристик.

Частная психодиагностика связана с конкретным использованием тестов в разных областях психологической практики. Например, профессиональная психодиагностика имеет дело с измерением психологических свойств субъекта в профессиональной деятельности, детская психодиагностика связана с использованием тестов для диагностики индивидуально-психологических особенностей детей, патопсихологическая диагностика использует тесты для выявления особенностей психики у людей, страдающих разного рода психическими заболеваниями[1].

Важной составной частью общей психодиагностики является ***психометрика*** – область психодиагностики, связанная с теорией и практикой измерения в психологии [4]. Выделяют ***общую психометрику и дифференциальную психометрику.***

Общая психометрика имеет дело с измерением психологических характеристик стимулов. Она предполагает моделирование присущих всем людям функциональных зависимостей между

---

[1] В данном учебном пособии мы ограничимся только рассмотрением вопросов, связанных с общей психодиагностикой. Поэтому в дальнейшем, употребляя термин «психодиагностика», мы будем иметь в виду общую психодиагностику.

стимулами и психологической реакцией на них. Эта область психометрики явилась исторически первой и нашла свое отражение в психофизике. В дальнейшем область психометрики чрезвычайно расширилась, что и привело, в свою очередь, к появлению дифференциальной психометрики.

Дифференциальная психометрика предполагает измерение не стимулов, а индивидов, их способностей, психических свойств, когнитивных функций (особенностей внимания, памяти, мышления и т.п.), а также особенностей эмоционально-мотивационной сферы [4]. Именно эта отрасль психометрики и является методологической основой психодиагностики, обосновывающей требования к измерительным процедурам, их разработке, проведению и интерпретации тестовых данных.

1.2. Основные задачи психодиагностики

Использование психодиагностических методов связано с выработкой и принятием обоснованных решений, которые касаются либо отдельной личности, либо группы людей. С этой точки зрения психодиагностика имеет отношение к решению двух типов задач: оценке педагогического метода и оценке личности [6].

Оценка педагогического метода преследует реализацию двух целей: с одной стороны, определение эффективности одинаковых воздействий, направленных на разных людей; а с другой стороны, сопоставление эффективности разных воздействий на одинаковые по своим психологическим особенностям личности.

Оценка личности предполагает определение наличного уровня выраженности ее психологических особенностей и процесса развития личности и зависит от сферы применения. Психологические тесты нашли широкое распространение практически во всех сферах работы практического психолога, однако наибольшее применение они приобрели в сфере образования, в клинической практике, а также в области профессиональной деятельности личности.

Использование тестов в сфере образования. В области образования тесты, прежде всего, используются при диагностике способностей и знаний учащихся, для проверки эффективности

обучающих программ. В настоящее время введение различных дифференцированных обучающих программ предполагает и разработку методов объективного контроля, которые бы на основании учета исходного уровня и возможностей развития позволили повысить эффективность всех видов обучения. Типичными для данного направления могут быть тесты способностей, тесты достижений и интеллекта.

Другим типом тестов, используемых в сфере образования, могут быть измерительные инструменты, предназначенные для отбора школьников при поступлении в высшие учебные заведения. Это, прежде всего, тесты достижений для разного рода дисциплин. Прообразами таких тестов являются тестовые задания, используемые при проведении Централизованного тестирования в нашей стране, результаты которого засчитываются как результаты вступительных экзаменов в большом количестве российских вузов. Студенты и школьники, желающие обучаться в образовательных учреждениях США, в обязательном порядке проходят стандартизированный тест TOIFIL, направленный на диагностику знаний английского языка.

Помимо тестов способностей в практике работы школьного психолога могут использоваться и тесты диагностики интересов школьников с целью выяснения предпочитаемой сферы будущей профессиональной деятельности и распределения их по профильным классам.

Использование тестов в клинической практике. Тесты и другие методы прикладной психологии (интервью, наблюдение, нейропсихологические пробы) широко используются в клинической практике для постановки диагноза. Ситуации, в которых у психолога возникает потребность в использовании психологических тестов, могут быть следующие:

– применение тестов для уточнения диагноза;

– оценка особенностей интеллекта, способностей и личностных особенностей детей, испытывающих трудности в адаптации к учебе, если они связаны с неврологическим дефицитом, сниженным умственным развитием, личностными проблемами или совокупным влиянием вышеизложенных факторов;

– при проведении психотерапевтической работы с пациентами как способа фиксации изменений, происходящих в его поведении и эмоциональных состояниях.

В зависимости от целей исследования в клинической практике могут быть использованы интеллектуальные тесты, личностные опросники, проективные методики, нейропсихологические пробы и другие специальные инструменты.

Использование тестов в сфере консультирования и коррекции поведения. Данная сфера несколько выпадает из предложенной классификации, поскольку консультирование может встречаться и в образовании, и в клинике, и в области бизнеса и профессиональной деятельности. Вместе с тем она имеет некоторые общие моменты, независимо от области применения: возможность корректировать поведение клиентов. В ходе консультирования психолог должен выявить проблему, возникшую у клиента, и причины ее появления, предложить способы воздействия на индивида с целью изменения его поведения и, наконец, предложить методы измерения степени эффективности воздействия на клиента. Для того чтобы проверить эффективность психокоррекционных приемов, психолог должен замерить некое исходное состояние субъекта, затем проводить замеры в ходе проведения психотерапии или психологических тренингов и потом произвести контрольный замер результата воздействия. Например, если консультант установил, что проблема трудностей с учебой у ребенка связана с низким уровнем коммуникабельности и отсутствием навыков общения, он может предложить провести с данным ребенком (или группой) коммуникативный тренинг. Но перед проведением данного тренинга он обязательно должен замерить наличный уровень выраженности коммуникативных качеств у данного конкретного ребенка и затем произвести аналогичные замеры после окончания тренинга.

Использование тестов в сфере бизнеса и профессиональной деятельности личности. Одной из основных прикладных проблем в данной сфере является диагностика уровня профессиональной пригодности субъекта. При решении этого вопроса психодиагностика исходит из решения двух задач:

– оптимальное решение проблемы выбора профессии человеком путем соотнесения его индивидуально-психологических особенностей и требований, предъявляемых к нему профессией;

– психологический профессиональный отбор наиболее подходящих кандидатов для выполнения конкретных профессиональных обязанностей.

Для оценки профессиональной пригодности могут использоваться тесты профессиональных способностей и интересов, опросники профессиональных предпочтений, личностные тесты и даже проективные методики.

Наш собственный опыт показывает, что хорошо зарекомендовали себя в этом плане батарея профессиональных способностей (DAT) [45], опросник профессиональных предпочтений Дж. Холланда [23], Дифференциально-диагностический опросник Е.А. Климова [37], Тест умственных способностей (ТУС), представляющий собой русскоязычную версию теста структуры интеллекта Р. Амтхауэра [43] и некоторые другие.

Помимо указанных сфер тесты и другие психодиагностические процедуры могут использоваться и в других областях прикладной психологии. В юридической психологии тесты являются одним из основных инструментов при проведении судебно-психологической экспертизы [14]. В психологии спорта тесты могут быть использованы при диагностике предстартовых состояний спортсменов, для определения степени совместимости спортсменов в командных видах спорта и т.п. [39].

Наконец, психологические тесты играют важную роль и в процессе построения психологической теории. Результаты тестирования могут быть использованы для подтверждения или опровержения выдвинутых гипотез относительно природы исследуемого явления, его структуры и связей с другими психологическими феноменами.

Вопросы для самопроверки

1. В чем сущность предмета общей психодиагностики?
2. Перечислите основные сферы использования психологических тестов.
3. Для каких целей используются тесты в сфере профессиональной деятельности и бизнеса?
4. Что такое психометрика?

Глава 2. **История развития психодиагностики**

История психодиагностики, как и любой науки, традиционно рассматривается в хронологическом порядке [1], [2], [49]. Другой путь ее анализа заключается в выявлении основных источников становления, которыми считаются экспериментальная психология, дифференциальная психология, тестология и прикладная психология [36]. Мы попытаемся предложить некий третий вариант, представляющий собой комбинацию вышеуказанных путей. Он будет представлять собой традиционную хронологию развития представлений о диагностике человека с выделением ряда краеугольных этапов в ее становлении.

Предыстория тестов уходит в глубину веков и связана с испытанием различных способностей, знаний, умений и навыков человека. Практически во всех странах **Древнего мира** (Египет, Китай, Греция, Вавилон) мы находим некие способы проверки различных умений и способностей человека для определения их профессиональной пригодности [1]. Нередко такие испытания становились предметом гордости своего народа, а иногда служили и для решения экономических проблем. Из древних источников известно, что индийский принц Девсарн, желая испытать мудрость иранцев, прислал им шахматы с предложением разгадать правила этой игры. В случае неудачи иранцы обязаны были выплатить Индии подать. Однако визирь Хосрова Важурмир понял правила игры и, в свою очередь, изобрел игру, послав ее индийцам с аналогичной просьбой. Индийцы так и не смогли разгадать смысл посланной игры, которая в настоящее время известна как нарды [7].

Историки традиционно связывают появление первых примитивных форм тестирования умений и знаний с Древним Китаем [59]. Установлено, что относительно изощренные программы

тестирования государственных служащих существовали более 4000 лет назад [59]. В течение трех тысяч лет в Китае существовала открытая и конкурсная система испытаний для оценки знаний и умений в различных областях, таких как музыка, стрельба из лука, верховая езда, умение читать и писать. Устные экзамены проводились в Китае каждые три года, и их результаты использовались для оценки эффективности деятельности и продвижения по службе. Во время правления Династии Хань (с 206 г. до н.э. по 220 г. н.э.) применялись тестовые батареи из двух или более тестов для оценки таких дисциплин, как гражданское право, военное дело, сельское хозяйство, финансы, география. Особенно эффективно система тестирования развивалась во времена правления Династии Минг (1368 – 1644 гг. н.э.) и сохранилась вплоть до 1905 года, когда имевшие место реформы ее разрушили. Экзамен проходил в три обязательных этапа и включал проведение тестирования, как на региональном, так и на федеральном уровне. Первый этап проводился во всех главных городах Империи. Претенденты, все лица мужского пола, располагались в тесных маленьких хижинах, сделанных специально для целей тестирования, где они сочиняли поэму и писали на заданную тему в течение дня и ночи. Число претендентов, выдержавших это испытание благодаря своему красноречию и каллиграфическому почерку, составляло всего от 1 до 7% от числа всех участников тестирования. Каждый третий год все мужчины, прошедшие первое испытание, проходили через вторую, более суровую стадию тестирования, на проведение которой отводилось три дня и три ночи. После этой стадии испытаний, которую проходили 1 – 3% претендентов, осуществлялась последняя, третья фаза тестирования, проходившая в столице Китая Пекине, после которой приблизительно 3% кандидатов могли быть допущены для государственной службы [59].

Появление тестирования на Западе произошло благодаря знакомству с Китаем. Отчеты миссионеров и дипломатов, работающих в Китае, показывают, что Британская Восточно-Индийская компания скопировала китайскую систему в 1832 году как метод для отбора служащих для работы за рубежом. Эффективность китайской системы была настолько высока, что бри-

танское правительство в 1855 году внедрило сходную систему тестирования для отбора государственных чиновников. Примеру британского правительства вскоре последовали правительства Германии и Франции, а в 1883 году и правительство США [59], [61].

Этот пример показывает, что уже за три тысячи лет в мире существовали системы испытаний, основанные на некоторых принципах психодиагностики, которые существуют и в настоящее время.

Созданные в древности умозрительные системы испытаний способностей и умений требовали определенной теоретической базы. Известный древнегреческий философ Платон (427 – 347 г. до н.э.) утверждал, что люди должны выбирать себе работу в соответствии со своими способностями и дарованиями. Клавдий Гален (131 – 205 г. н.э.) проводил многочисленные опыты, пытаясь доказать, что вместилищем интеллекта является мозг, а не сердце, как было принято в то время. Интересно, что начало «Темных Веков» в истории связывают со временем смерти Галена, когда интерес к личности, медицине и философии уступил место религии.

В Средние века господствовавшие в обществе религиозные представления, нашли свое отражение и в понимании психологии людей. В частности в трудах таких известных теологов и философов Средневековья, как Святой Августин (354 – 430), Фома Аквинский (1225 – 1274), все психические проявления человека (восприятие, мышление, самоконтроль) объявлялись делом рук Творца.

В обществе, основанном на догматах церкви, когда все естественные катастрофы объявлялись делом рук Сатаны, люди, отличающиеся от средней массы своими взглядами, поведением, объявлялись существами, общающимися с дьяволом. Главный вопрос того времени «Кто является одержимым дьяволом?» был зафиксирован в булле папы римского, написанной в 1484 году, в которой глава католической церкви просил людей о помощи в распознавании слуг дьявола. Своеобразным откликом на это обращение был написанный двумя доминиканскими монахами Генрихом Крамером и Иоханом Шпрангером трактат под названием

«Malleus Malificarum»», что в переводе означает «Молот ведьм». В первых двух частях этого документа декларируется положение о существовании ведьм и колдунов, подкрепленное детальными доносами о фактах столкновения с дьяволом. С точки зрения нашего предмета наиболее интересной является третья, заключительная часть, в которой приводятся способы распознавания людей одержимых дьяволом. Например, чтобы определить, является ли женщина ведьмой, ее раздевали донага и, связав руки и ноги, бросали в реку. Если женщина тонула, то вопроса о ее дьявольской сущности не возникало. Если же она продолжала плыть, то это доказывало, что она в самом деле является ведьмой, и ее ждали пытки и сожжение на костре [52].

Идеи диагностики способностей людей в феодальном обществе были подчинены практическим целям, и основным заказчиком выступала армия. Такая ситуация была характерна и для России той эпохи. Одним из первых документов в этом плане явилось «Соборное уложение царя Алексея Михайловича», написанное в 1649 году. В нем указывалось на необходимость осмотра людей, поступающих на государеву службу, который должны проводить лекари, оценивающие здоровье и физические способности людей. И хотя речь в основном шла о медицинской диагностике, но уже в ряде медицинских руководств того времени было записано, что если лекарь не найдет никаких изъянов в физическом здоровье рекрута, он должен приступить к оценке его душевных способностей [27].

Подводя краткий итог этому периоду развития, необходимо отметить, что для существования психологической диагностики не было ни экономической, ни научной базы. И если медицина накопила к тому времени определенный опыт диагностики физического здоровья, то при оценке психических качеств, практики опирались на свой субъективный опыт и существовавшие в обществе стереотипы и штампы при оценке личности.

Появление в обществе капиталистических отношений, конкурентной борьбы заставило предпринимателей искать новые пути интенсификации производства, в том числе и связанные с отбором кадров, предполагавшим оценку их психических способностей, знаний, умений и навыков. Однако психологическая

наука, находившаяся в недрах философии, не могла быть методологической базой, способной обеспечить решение выдвинутых практикой проблем. Требовался принципиально иной подход, который стал возможен с появлением **экспериментальной психологии.**

Возникновение экспериментальной психологии связывают с именем немецкого ученого **Густава Фехнера** (1801 – 1887), впервые обосновавшего психофизический метод исследования. Для описания зависимости между интенсивностью ощущения и силой раздражителя Г. Фехнер использовал логарифмическую функцию, что позволило ему в 1860 году сформулировать первый количественный закон в психологии. Заслуга Г. Фехнера заключается в том, что он реально показал возможность измерения психических явлений по аналогии с физическими, что впоследствии было взято на вооружение психодиагностикой в качестве методологического принципа.

Опираясь на простейшие опыты Фехнера, немецкий психолог **Вильгельм Вундт** (1832 – 1920) в 1870 году основал первую экспериментальную лабораторию по изучению психических свойств, заложив тем самым основы психологии как экспериментальной науки. Экспериментальная психология занималась изучением простейших психических процессов и свойств, возникающих под влиянием определенных раздражителей, с целью раскрытия некоторых общих закономерностей, которые она пыталась представить по возможности в форме математических моделей. Эксперименты В. Вундта по исследованию индивидуальных различий, касающихся времени простой сенсомоторной реакции, привели к созданию комплекса «личных уравнений» этой простейшей психической функции. В дальнейшем было обнаружено, что аналогичные уравнения можно составить и для других психических функций. Стимулом для возникновения психодиагностики послужил и предложенный экспериментальной психологией большой арсенал аппаратурных методов [6], [49].

Вместе с тем не совсем верным является тезис о том, что психодиагностика непосредственно выросла из экспериментальной психологии. Доказательством этого является убежденность В. Вундта в невозможности экспериментального изучения выс-

ших психических функций, составляющих сущность личности [6].

Однако естественно-научный подход к исследованию психики и тот стиль мышления, который сложился в школе В. Вундта, определили дальнейшее развитие психодиагностики. Исследования индивидуальных различий в школе Вундта подготовили почву и для развития **дифференциальной психологии,** ставшей важным источником в развитии психодиагностики.

Возникновение дифференциальной психологии как отрасли психологии изучающей индивидуальные различия, было подготовлено в 1896 году докладом **Альфреда Бине и Виктора Анри** «Индивидуальная психология», в котором авторы указывали на возможность изучения высших психических функций (памяти, мышления, представлений) с помощью экспериментального метода, что противоречило господствовавшей точке зрения, декларируемой В. Вундтом.

Появление психологии индивидуальных различий часто связывают с именем выдающегося английского ученого сэра **Фрэнсиса Гальтона** (1822 – 1911), бывшего одной из влиятельных фигур в истории психологического тестирования. Будучи родственником Чарлза Дарвина, Гальтон был заинтригован его работами в области теории эволюции. Изучая наследственность на примере традиционных биологических объектов (горох) Ф. Гальтон постепенно менял свои интересы в сторону изучения наследственности человека и различных путей измерения человеческих способностей. Он был сторонником массовых обследований, поскольку считал, что только на этой основе можно понять единичное и особенное. На выставке в Лондоне в 1884 году им демонстрировалась антропологическая лаборатория, где за символическую плату (3 – 4 пенса) любой человек мог измерить ряд своих характеристик как антропометрических (рост, вес), так и психофизических (различения цветов, память на форму и т.п.). Всего им было обследовано 9377 человек по 17 показателям [1], [49]. Свои методы изучения он называл тестами.

Важной заслугой Ф. Гальтона было развитие идей корреляционного и регрессионного анализа. Еще в своих опытах над горохом он использовал понятие коэффициента корреляции, кото-

рое является центральным понятием в экспериментальной психологии и психодиагностике и которое в дальнейшем было доведено до технологического уровня в математической статистике. При сопоставлении роста родителей и детей им была открыта тенденция, заключающаяся в стремлении потомства к среднему росту, которая была объявлена «всеобщим законом природы» и названа «регрессией к среднему» [1].

Ф. Гальтон был не только первым создателем тестов для диагностики ума, но и пионером в области изучения личностных особенностей. При этом он отмечал, что точное измерение личностных черт человека может дать статистика его поведения в повседневных ситуациях. Кроме того, Ф. Гальтон стоял и у истоков проективной психологии. Проведя серию ассоциативных экспериментов, он пришел к совершенно правильным выводам о природе ассоциаций, их свойствах и проявлениях. В частности, он рассматривал ассоциации как глубоко индивидуальные образования, которые теснейшим образом связаны со сферой бессознательного, причем наиболее стойкие ассоциации связаны с детством. И хотя его исследования в этом направлении остались незавершенными, именно он заложил основы той техники, которая спустя почти полвека получила название проективной.

Хотя Ф. Гальтон внес значительный вклад в изучение психологии индивидуальных различий, однако подлинным ее зачинателем считают немецкого психолога **Вильяма Штерна** (1871 – 1938), опубликовавшего в 1900 году работу «О психологии индивидуальных различий», которая привела к широкой разработке проблем дифференциальной психологии, приобретшей с этого момента права гражданства [6], [49]. Вместе с тем, в работах В. Штерна речь шла не о различиях между людьми, а о различиях в психических функциях. Проводились сравнительные исследования разных популяций, в ходе которых выявлялись возрастные, половые и т.п. различия между группами. Следовательно, методологически дифференциальная психология ничем не отличалась от экспериментальной психологии, поскольку изучалась не конкретная личность, а отдельные психические свойства, взятые изолированно друг от друга. Возникновение же психодиагностики характеризовалось переходом от исследования индивидуаль-

ных различий в психических функциях к исследованию индивидуальностей. В этом и кроются, по мнению Г. Витцлака, различия между дифференциальной психологией и психодиагностикой. Дифференциальная психология изучала общие закономерности различий между людьми, и основным ее результатом является определение типологии психических функций, процессов и состояний. Психодиагностику же интересует конкретная личность, а не усредненный субъект [6].

Идеи Ф. Гальтона и конкретное их воплощение в виде экспериментальных процедур были подхвачены рядом психологов. Наиболее известным их проводником в США стал **Джеймс Маккин Кеттелл** (1860 - 1944), которого считают основателем тестологии. Дж. Кеттелл был учеником Вундта и после защиты докторской диссертации в Лейпциге вернулся в США, где работал в Пенсильванском университете. Затем он вернулся в Европу, где преподавал в Кембриджском университете и познакомился с Ф. Гальтоном, которого описал «как величайшего человека, которого он когда-либо знал» [67, с. 96]. В 1890 году, после возвращения в Пенсильванию, им была написана статья «Умственные тесты и измерения», оказавшая решающее влияние на разработку точных методов диагностики. Им был введен термин «умственный тест», кроме этого, он называл тестами методы анализа индивидуальных различий, созданные другими психологами. Предложенный Д. Кеттеллом список «тестов» предполагал оценку различных физических и психических способностей. Представление Д. Кеттелла о тесте существенно отличалось от методов, которые использовались в предшествующие эпохи. Он определял тест как экспериментальную процедуру, которая должна отвечать всем требованиям, предъявляемым к эксперименту: одинаковые условия проведения для всех испытуемых, ограничение во времени в пределах часа, отсутствие посторонних и т.п. Результаты тестирования должны подвергаться статистическому анализу, путем вычисления средних значений, среднего отклонения и т.д.

Работы Д. Кеттелла существенно продвинули методологию тестирования, уточнив содержание теста как экспериментальной процедуры и включив требования стандартности условий. Более

того, в работах Д. Кеттелла была высказана мысль о сравнении результатов тестирования с каким-либо внешним, независимым от эксперимента критерием для определения информативности теста, т.е. впервые прозвучала одна из основных идей современной психодиагностики – идея внешней валидности. Д. Кеттелл, в свою очередь, высказал предположение о том, что повторение теста будет одним из лучших критериев проверки его пригодности [1].

Таким образом, к началу XX века и в течение 10 - 20-х годов в психологии сложилось понимание теста как экспериментальной процедуры и выдвинулись две основополагающих концепции оценки качества теста: надежности и валидности. Задача заключалась в реализации этих концепций и доведении их до уровня конкретной технологии, что было достигнуто путем проникновения в психологию методов **математической статистики.**

Важным шагом на этом пути стал предложенный английским математиком **Карлом Пирсоном** (1857 - 1936) коэффициент корреляции и целый ряд других показателей как меры количественной связи между фактами. Теория корреляции стала первой теорией, ставшей основой для создания теории тестов и существенно изменила всю методологию их конструирования, обоснования и применения.

Заметный вклад в теорию тестирования внес другой английский психолог, ученик В. Вундта **Чарлз Спирмен** (1863 - 1945). Он попытался соединить теорию физических измерений с психологией и корреляционным методом. В результате этого им были предложены методы для оценки надежности теста, а также предложен собственный коэффициент корреляции, пригодный для малых выборок и основанный на ранжировании. Этот коэффициент получил в математической статистике название «коэффициента ранговой корреляции» или «коэффициента корреляции Спирмена». Разработанный Спирменом метод однофакторного анализа внес существенный вклад в развитие методологии психометрической проверки тестов и исследования интеллекта.

Дальнейшее применение методов математической статистики в психодиагностике связано с именем американского социолога **Луиса Терстоуна** (1887 - 1955). Терстоун пришел в социологию

из психиатрии и психофизиологии и попытался применить технику психометрического шкалирования в социологических исследованиях. Созданные им тесты и шкалы до сих пор используются в современных психологических и социологических исследованиях. Кроме того, им была усовершенствована методика факторного анализа Ч. Спирмена и предложен метод многофакторного анализа, который использовался им для эмпирических исследований интеллекта.

Проникновение аппарата математической статистики, и прежде всего корреляционного анализа, в психодиагностику существенно изменило представления психологов о методологии тестирования, что выразилось в попытке ее конкретной реализации в практике тестирования.

В 1905 году французский психолог Альфред Бине (1857 – 1911) вместе со своим коллегой Теодором Симоном сделали попытку построить первый тест умственного развития. Их новаторство заключалось в том, что они осознали необходимость проверки теста еще до его практического использования. Созданный ими тест состоял из 30 заданий, каждое из которых было отобрано после тщательной проверки до применения на практике. В качестве критериев использовалась мера трудности задания для детей различного возраста и степень совпадения тестовых результатов с оценками преподавателей. Использование стандартной инструкции было призвано минимизировать субъективизм при проведении теста разными исследователями.

Вслед за шкалой умственного развития Бине – Симона разработка тестов и их проникновение в практику стало осуществляться быстрыми темпами. Немалую роль в этом сыграла и первая мировая война, потребовавшая проведения мобилизации в кратчайшие сроки и быстрого определения профессиональной пригодности призывников.

После окончания войны тесты начали быстро распространяться во многих странах мира, в том числе и в России. Тест зарекомендовал себя хорошим инструментом для выявления индивидуальных различий в различных областях практики. Наиболее интенсивно тестовое направление стало развиваться в 20 – 30-е годы, особенно в педагогической сфере и промышленности. К

1928 году имелось примерно 1300 тестов, по которым в течение одного года было получено около 30 миллионов тестовых результатов [1].

Развитие тестирования в России, начавшееся в 20-е и 30-е годы, было связано, прежде всего, с бурным развитием двух направлений практической психологии: психотехникой и педологией. Оба эти направления предполагали изучение и диагностику индивидуальных различий в трудовой деятельности (психотехника) и в образовании (педология). Поскольку оба эти направления как направления прикладной психологии были призваны выявить наиболее успешных работников и наиболее одаренных учеников, то тестированию, прежде всего, подвергались интеллект и специальные способности личности. Диапазон применяемых тестов был довольно широким. Применялись десятки разновидностей тестов. Помимо классических тестов А. Бине и их разнообразных модификаций использовались «профили» Россолимо, тесты Меде и Пиортковского, тесты Левитова и многие другие. В психотехнике употреблялись тесты профессиональной пригодности, в педологии – тесты возрастного развития, тесты для отбора «высокоодаренных» детей, в патопсихологии – «конституционные» тесты. Тестирование личностных особенностей практически не проводилось, за исключением диагностики развития моральных качеств, где использовался «метод коллизий», представляющий собой разновидность «тестов морального развития».

Пропаганда и теоретическое обоснование метода тестов в России связаны были с деятельностью таких психологов, как Г.И. Россолимо, М.А. Бернштейн, П.П. Блонский и другие. Активность тестологов особенно усилилась в период I педологической конференции (апрель 1927), где развернулась дискуссия о тестах и их значении. В мае того же года было организовано объединение московских тестологов, и намечены широкие перспективы работы. Практическая работа по тестированию осуществлялась многочисленными педологическими кабинетами отдельных детских клиник, школьных лабораторий, кабинетами психотехники и т.п. Издавались многочисленные сборники тестов [26].

Тенденции развития советской психологии того времени заключались в акцентировании внимания психологов на исследо-

вании прежде всего психических процессов (ощущений и восприятия, внимания, памяти, мышления), психомоторных функций, свойств темперамента. Вопросы изучения личности, ее структуры, мотивации поднимались на высоком теоретическом уровне, не доведенном до уровня практической реализации. В этой связи тестированию в основном подвергались свойства психических процессов (внимания, памяти), отдельные мыслительные операции, психомоторные характеристики (координация движений обеих рук, тремор, сенсомоторная координация).

Однако активное внедрение прикладной психологии в дело переустройства общества столкнулось с «чистой», академической психологией. Такое размежевание особенно сильно было заметно в конце 20-х – начале 30-х годов. Оба направления обвиняли друг друга во «всех смертных грехах». Представители прикладного направления критиковали психологию за академизм, отрыв от жизни, неспособность воспринимать новое. В свою очередь академические психологи обвиняли оппонентов в голом практицизме, отрыве от теории, чрезмерном увлечении тестами как единственным методом. Наметившийся кризис был характерен не только для нашей страны, но и для других стран мира. В частности, в США противники тестов привлекали внимание общественности к тестам как средству вторжения в частную жизнь граждан и нарушению этических принципов. Но если в западных странах критика тестов не выходила за рамки полемики в прессе и публичных выступлений, то в нашей стране дело приняло несколько иной оборот.

В середине тридцатых годов дискуссия относительно педологии и психотехники переросла в практические действия по отношению к этим направлениям прикладной психологии. Особенно шумной была кампания против педологии. Её вершиной стало печально известное постановление ЦК партии большевиков «О педологических извращениях в системе Наркомпроса»[2], принятое в 1936 году. В этом документе педология признавалась лженаукой, классово-враждебным течением, издевательски относящимся

---

[2] Народный комиссариат просвещения – название Министерства образования в СССР в 30-е годы.

к детям и наносящим вред советскому государству. Разгром педологии был вызван рядом причин:

1. В то время завершился разгром «старых большевиков» – последователей В.И. Ленина (Н.И. Бухарин, А.А. Бубнов, Н.К. Крупская), а также Л.Д. Троцкого и его учения, которые поддерживали педологическое и психотехническое направления. Кроме того, педология не выполнила своей основной миссии – не дала научное обоснование идеологии большевизма. В частности, не соответствовали принципу партийности полученные результаты о более низком интеллектуальном уровне советских школьников по сравнению с их американскими сверстниками, об отставании в умственном развитии детей из семей рабочих и крестьян от детей интеллигенции, данные о национальных различиях в коэффициенте интеллекта.

2. Определенную роль в судьбе педологии сыграл и тот факт, что, помимо квалифицированных специалистов, тесты попадали в руки дилетантов. Открытие педологических лабораторий распахнуло двери большому количеству некомпетентных в психологии людей (несостоявшихся чиновников, учителей, врачей, инженеров), которые считали, что должность психолога не требует специальной подготовки. Это компрометировало науку в глазах основных заказчиков тестирования и провоцировало расправу над ней [36], [49].

Непосредственное вмешательство идеологии в науку привело к закрытию всех центров и лабораторий, печатных изданий, ликвидации ученых степеней, арестам, ссылкам и физической расправе над ведущими специалистами в области прикладной психологии. Тест как инструмент диагностики исчез из отечественной психологии.

Начало второй мировой войны вновь привлекло внимание к тестированию. Тест оказался наиболее экономичным и надежным методом для диагностики профессиональной пригодности призывников и расстановки их по конкретным специальностям и должностям. Особо острую необходимость в этом испытывали технически сложные виды Вооруженных сил (авиация, военно-морской флот). Наиболее интенсивно взялись за развитие армейского тестирования американцы после официального вступления

США в войну, когда японская авиация разбомбила американский флот в бухте Перл-Харбор.

Ими было предложено и реализовано два крупных проекта [1].

Первый был связан с адаптацией призывников к условиям боевых действий и снижением чувства страха. Реализация этого проекта требовала исследования причин возникновения страхов, выявления типов личности, наиболее устойчивых к боевым стрессам. С этой целью необходимо было использование надежных и валидных тестов.

Второй проект касался профессионального отбора летчиков. Большинство профессий авиационного профиля требовало диагностики самых разнообразных качеств: интеллекта, личностных черт, психомоторных способностей, физического здоровья. Использование тестов в этом проекте доказало их высокую эффективность. Применение тестов при профотборе летчиков позволило примерно в 2,5 раза сократить отсев из училищ по непригодности. По мнению известного американского психолога Д. Гилфорда, каждый доллар, затраченный на разработку тестов, позволил сэкономить тысячу долларов [1].

После окончания второй мировой войны интерес к тестам на Западе заметно упал, что было, видимо, связано с отсутствием необходимости широко использовать армейские тестовые батареи.

Интерес к тестированию возник вновь в конце 50-х годов после запуска в Советском Союзе первого искусственного спутника Земли. Успехи советской космонавтики стимулировали американских ученых для анализа причин этой ситуации, в результате которого было установлено, что Советское государство уделяет особое внимание развитию системы образования, особенно в области техники и естественных наук. Это заставило американцев обратить самое пристальное внимание на образование в собственной стране, что нашло свое отражение в Национальной программе защиты образования, принятой Конгрессом США. Реализация этой программы потребовала и колоссальной переработки старых тестовых заданий по всей стране. В ходе проведения пи-

лотажного обследования на выборке 140 000 учащихся использовались тесты различных способностей.

Акцентирование внимания на использовании тестов в различных сферах практики давало не только положительные результаты, но и породило ряд проблем.

Во-первых, успешное использование тестов в годы войны создало прочный базис применения их в психологической практике и в мирное время. Однако по мере расширения масштабов тестирования в американском обществе вновь возникло беспокойство по поводу бесцеремонного вторжения тестов в личную жизнь, сокрытия результатов тестирования и т.д. В результате большой процент американцев выступил против использования тестов в практике профессионального отбора и продвижения по служебной лестнице, а также в школьной практике.

Кроме того, возник ряд проблем, связанных с технологией конструирования и проверки тестов и этическими нормами их применения.

Решение этих проблем в американской психологии привело к созданию Стандарта по конструированию и использованию тестов, в котором рассматривались основные принципы и требования к созданию тестов, а также нормы профессиональной этики для создателей и пользователей тестов. Этот Стандарт был утвержден Американской психологической ассоциацией и является юридическим актом, регламентирующим работу психолога в области тестирования [68]. Например, для приобретения теста в США необходимо обязательно давать сведения о наличии у клиента соответствующего документа об образовании и лицензии на право заниматься практической деятельностью. Аналогичные нормативные акты существуют и в европейских странах.

В настоящее время производством тестов за рубежом занимаются специальные фирмы, каждая из которых специализируется на определенных их видах. В США известны три крупные фирмы, занимающиеся производством тестов: Psychological Assessment Resources (PAR), Consulting Psychologists Press (CPP) и Western Psychological Services (WPS).

Издательская компания ***PAR,*** центр которой расположен в штате Флорида в г. Одесса, специализируется на стандартных

личностных опросниках, таких как Миннесотский многофакторный личностный опросник (MMPI), личностный опросник NEO-PI, методиках диагностики профессионального стресса, а также ряде проективных методик (тест Роршаха).

Калифорнийская компания ***Consulting Psychologists Press*** известна публикациями шкал установок, шкал рабочей среды, а также выпуском таких известных методик, как опросник на диагностику интересов Э. Стронга, Калифорнийский личностный опросник, опросник на диагностику межличностных отношений В. Шутца (FIRO), опросник диагностики психического выгорания К. Маслак (MBI).

Продукцией ***Western Psychological Services,*** располагающейся в Лос Анжелесе, являются в основном проективные методики (тест тематической апперцепции (ТАТ) и его разновидности, рисуночные тесты, тест «Рука»), а также стандартные личностные опросники MMPI и 16PF Р. Кеттелла и некоторые интеллектуальные тесты

В Европе существует сеть крупных компаний, работающих в области психологического тестирования, объединившихся в 1991 году в «Европейскую группу издателей тестов». Основными задачами этой ассоциации является работа в области разработки тестов в соответствии с международными стандартами, обеспечения пользователей высококачественной продукцией и подготовка специалистов в области тестирования. В нее входят компании практически всех ведущих европейских стран: Англии, Германии, Франции, Голландии, Италии, Испании, скандинавских стран [4].

В нашей стране работа по психодиагностике, начиная с конца 30-х годов и до начала 60-х, практически не проводилась. Период «оттепели» во времена правления Н.С. Хрущева начал процесс освобождения психологии в целом, и прежде всего психологии личности, от идеологических рамок, что получило свое дальнейшее развитие в середине 80-х годов с началом «перестройки». Это проявилось в исследовании реальной личности, в переходе от глобальных концепций к анализу конкретных ее аспектов (мотивов, личностных характеристик), к рассмотрению личности с позиций гуманистического подхода. Гуманистический подход к

личности проявился не только в приложении идей зарубежной гуманистической психологии на российскую почву, но и в приложении стратегий и тактик зарубежной психодиагностики. Отечественные психологи снова стали активно использовать в своем арсенале психологические тесты как основной инструмент диагностики личностных черт. Наибольшую популярность имели личностные опросники Г. Айзенка и 16 PF Р. Кеттелла, а также опросник MMPI, первая русскоязычная версия которого была создана Ф.Б. Березиным в 1967 году. Были созданы также версии московскими психологами (Л.Н. Собчик, 1971) и в Санкт-Петербургском психоневрологическом институте им. В.М. Бехтерева. Попытки российских психологов сконструировать собственные типологии личности неизбежно приводили их к созданию собственных личностных опросников. Так, А.Е. Личко в 1970 году был создан личностный опросник для определения типов акцентуаций и вариантов конституциональных психопатий в подростковом и юношеском возрасте (14 – 18 лет). Он диагностирует 11 «чистых» типов акцентуаций и психопатий и ряд смешанных.

Помимо адаптаций личностных опросников появились и русскоязычные версии тестов интеллекта. В частности, известный тест структуры интеллекта Д. Векслера был адаптирован в Ленинградском психоневрологическом научно-исследовательском институте им. В.М. Бехтерева в 1969 году, а его детский вариант – московским психологом А.Ю. Панасюком в 1973 году [4].

Использование зарубежных тестов и попытки создания собственных тестов привели к появлению группы психологов в разных концах тогда еще Советского Союза, серьезно занимающихся проблемами конструирования, проверки и использования тестов. Это К.М. Гуревич и его сотрудники, В.Н. Дружинин, В.М. Русалов, А.Г. Шмелев – в Москве; В.К. Гайда, В.П. Захаров, А.Е. Личко – в Санкт-Петербурге; Л.Ф. Бурлачук – в Киеве. В России стали появляться и предприятия, занимающиеся распространением и продажей тестов. Особенно много их появилось в конце 80-х годов в связи с либерализацией законодательства в отношении частного предпринимательства. Большинство из этих фирм в основном занимались продажей зарубежных методик, по-

лученных «пиратским» способом, качество предлагаемой продукции было очень низким. Это привело, в конечном итоге, к прекращению деятельности такого рода предприятий. В настоящее время существуют три наиболее известные фирмы, специализирующиеся на разработке и распространении тестов: фирма «Иматон» в Санкт-Петербурге, Научно-производственный центр «Психодиагностика» в Ярославле и издательство «Смысл» в Москве.

Широкое вторжение современных компьютерных технологий во все сферы жизни российского общества не обошло стороной и психодиагностику. Использование компьютерных вариантов известных тестов, создание компьютерных психодиагностических систем, несомненно, способствует повышению эффективности как научных психологических исследований, так и практической работы специалистов-психологов [3], [47]. Разработкой и распространением компьютеризированных тестов занимается московская лаборатория компьютерной психодиагностики «Гуманитарные технологии» под руководством А.Г. Шмелева [47].

Бурное развитие тестирования в последние годы в России не только повысило эффективность работы практических психологов, но и породило ряд проблем. Прежде всего, значительная часть имеющих хождение в России личностных опросников не соответствует международным стандартам, предъявляемым к созданию тестов, что выражается в отсутствии психометрической проверки и отечественных норм. Эта тенденция усиливается и низкой квалификацией пользователей, слабо владеющих теорией и методикой конструирования тестов, широким и часто бесконтрольным распространением тестов и их публикацией в печати. Такая ситуация во многом вызвана отсутствием в отечественной психологии законодательной базы в отношении распространения и использования тестов, аналогичной той, которая существует в зарубежных странах. Имеющиеся в психологической практике варианты стандарта требований к психологическим тестам и норм профессиональной этики для разработчиков и пользователей тестов не имеют официального статуса [15], [19], [32].

Вопросы для самопроверки

1. Назовите научные дисциплины, ставшие основными источниками развития психодиагностики?
2. Дайте краткую характеристику особенностей развития психодиагностики в докапиталистическую эпоху.
3. В чем основная заслуга Ф. Гальтона в развитии психодиагностики?
4. Каковы особенности развития психодиагностики в нашей стране?

Глава 3. **Теоретические основы психодиагностики**

Основная проблема психодиагностики заключается в ответе на вопрос о том, что мы должны диагностировать, исходя из вектора «личность – ситуация». Триада «личность – ситуация – поведение» может развиваться в трех вариантах:

1. Одна и та же личность в разных ситуациях может демонстрировать одинаковое поведение.
2. Разные люди в одной и той же ситуации ведут себя одинаково.
3. Разные люди в одной и той же ситуации ведут себя по-разному.

Первый вариант объясняет данный феномен наличием личностных особенностей, которые проявляются во всех ситуациях. Например, агрессивные люди будут демонстрировать соответствующее поведение в профессиональных ситуациях, во взаимоотношениях с родными, в спорте, в других ситуациях повседневной жизни.

Тождественность поведения людей во втором варианте предполагает наличие таких особенностей ситуации, которые нивелируют специфику личности. Пожар в здании может вызвать паническую активность у всех находящихся в здании людей, независимо от их личностных особенностей.

Третий вариант предполагает, что одна и та же ситуация разными людьми может восприниматься по-разному, что и порождает различное поведение. Ситуация стресса может восприниматься тревожными людьми как угрожающая и вызвать нарушения в осуществлении профессиональных обязанностей, в то время как в сознании эмоционально стабильных людей данная ситуация может быть представлена как нейтральная и поэтому не вызовет дестабилизации профессиональной деятельности.

Таким образом, самый первый анализ проблемы «что мы должны диагностировать? » дает несколько альтернатив:

1. В соответствии с первым вариантом необходимо диагностировать стойкие личностные особенности;

2. В соответствии со вторым вариантом мы должны диагностировать ситуации;

3. Согласно третьему варианту нужно выявлять те когнитивные схемы, которые складываются в сознании человека.

Указанные альтернативы реализуются в трех основных направлениях психодиагностики: теории личностных диспозиций, ситуационизме и интеракционизме.

3.1. Теория личностных диспозиций

Теория личностных диспозиций исходит из положения о наличии стойких личностных особенностей, которые определяют поведение субъекта в различных ситуациях. В истории психодиагностики эта концепция конкретизируется в двух основных теоретических подходах: теории черт и теории типов.

Теория черт. Представители этого подхода считают, что любую личностную характеристику (тревожность, мотивацию достижений) можно представить как точку на некоем континууме от самого низкого значения ее выраженности до самого высокого значения. При этом оценки на этом континууме будут подчиняться нормальному распределению, т.е. крайнюю степень выраженности черты имеет небольшое количество людей, а большинство людей находятся в среднем диапазоне выраженности. Основателем данного подхода считают американского психолога Гордона Олпорта (1897 – 1967). Он ввел понятие черты, показав ее специфику по сравнению с другими единицами анализа, принятыми в психологии личности, и предложил классификацию черт.

Под чертой Г. Олпорт понимал некую нейропсихическую подструктуру, предрасполагающую человека вести себя определенным образом в широком диапазоне ситуаций [50]. С точки зрения концепции Олпорта поведение человека «надситуативно» во времени и пространстве. Надситуативность во времени предполагает наличие некоторого стабильного уровня выраженности

какой-либо черты (например, тревожности) в течение длительного периода времени, несмотря на присущие каждой личности спады и подъемы.

Надситуативность черты в пространстве предполагает постоянство поведения в широком диапазоне ситуаций, при этом речь идет о некотором стабильном среднем уровне выраженности черты для большинства ситуаций.

Г. Олпорт разделил все черты на кардинальные, центральные и вторичные.

Под кардинальной чертой понимается некая диспозиция, которая пронизывает все поведение человека таким образом, что в каждом его действии и поступке можно видеть ее влияние. Обычно у человека таких кардинальных черт немного, либо они вообще отсутствуют. Примером наличия кардинальных черт являются исторические фигуры или литературные персонажи, чье поведение было обусловлено этой чертой, которая становилась нарицательной. Например, макиавеллизм как безграничное стремление к власти – кардинальная черта известного правителя эпохи Возрождения Николо Макиавелли, ставшая нарицательным названием.

Центральные черты представляют собой диспозиции, которые проявляются в ограниченном круге ситуаций, однако их легко обнаружить, и они составляют базис личности. С точки зрения Олпорта, таких черт может быть от 3-х до 10.

Наконец вторичные черты наименее обобщены и заметны, и необходимо очень хорошо знать человека, чтобы их обнаружить. Обычно этот вид черт представляет собой различные установки индивида по отношению к конкретным ситуациям.

Г. Олпорт был поклонником идеографического подхода[3], заключающегося в описании личности как уникальной целостности. Данный подход он проиллюстрировал на примере работы,

[3] Понятие идеографического и альтернативного ему номотетического подхода введено немецким философом В. Виндельбандом для обозначения двух групп наук. Идеографический подход к познанию предполагал изучение уникального и индивидуального, в то время как номотетический подход был ориентирован на познание общих и универсальных законов.

получившей название «Письма Дженни» (1965), которая представляла собой 172 письма, посланных пожилой женщиной Дженни Мастерс в течение 12 лет. На основании этих писем он дал качественную и количественную характеристику ее личности, выделив центральные черты.

Идеи Олпорта об уникальности личности привели его, в конце концов, к мысли о наличии таких уникальных черт, которые недоступны научному познанию, что было проинтерпретировано как ненужность и невозможность существования науки об изучении личности. Это породило противоречие, существенно затруднявшее развитие психодиагностики. Однако введение им понятия черты и определение структуры личностных черт дало толчок для развития *номотетического* подхода к изучению личности, заключающегося в наличии общих для всех людей личностных черт, образующих иерархическую структуру. Данное направление получило свое развитие благодаря использованию факторного анализа, что позволило осуществить количественное измерение черт.

Одним из выдающихся представителей этого направления был известный американский психолог Раймонд Б. Кеттелл (1903 – 1998). Работа Р. Кеттелла заслуживает особого внимания, поскольку он предпринял попытку создать концепцию структуры личности на объективных основаниях. Для сохранения научной объективности он обратился к языку, выделив слова, характеризующие качества личности и поведения человека. На этом первичном материале он стал строить свою концепцию, опираясь на два метода: семантико-идеографические работы Г. Олпорта и метод факторного анализа. На основе факторизации им были выделены разные сочетания факторов, среди которых наиболее известна 16-факторная классификация. Согласно Кеттеллу, существует 16 первичных личностных черт и 4 вторичных фактора, по которым можно оценивать личность: замкнутость – открытость, серьезность – легкомысленность, покорность – напористость, робость – смелость, зависимость от группы – самодостаточность и т.д. [52].

Главным результатом проделанной Кеттеллом работы стал многофакторный личностный опросник 16 PF [51] , хорошо из-

вестный во всем мире, в том числе используемый и в нашей стране [19], [29], [47].

Другим крупным представителем номотетического подхода является известный английский психолог Ганс Юрген Айзенк (1916 – 1997). Г. Айзенк также использовал технику факторного анализа для измерения и классификации черт, но, в отличие от Кеттелла, он выделил два базовых параметра (суперфактора), которыми стали экстраверсия – интроверсия и нейротизм (эмоциональная стабильность – нестабильность). Позднее им был добавлен третий параметр, названный психотизмом. Люди, имеющие высокие баллы по этой характеристике, склонны к одиночеству, бесчувственны и невнимательны к чувствам других людей, а также могут находиться в оппозиции к общепринятым нормам поведения [61]. Айзенком было разработано несколько личностных опросников, построенных на его теоретической типологии, среди которых наиболее известен Личностный опросник Айзенка (Eysenk Personality Questionnaire) [4], [41].

Подводя итог, можно отметить, что, несмотря на принадлежность к одному теоретическому подходу в изучении личности – теории черт, имеются существенные различия в выделении и количества черт, и их качественного описания между представителями данного подхода. Если Г. Айзенк выделял только три обобщенных параметра черт, то Р. Кеттелл акцентировал внимание на почти 20 факторах. А предположение Г. Олпорта о наличии уникальных черт делало данный диапазон практически бесконечным. В связи с этим умы психологов занимал единственный вопрос: какое количество базовых черт достаточно для описания личности и есть ли модель, которая могла бы устроить большинство представителей теории черт?

В последнее время на роль такой модели претендует пятифакторная модель личности, или, как ее называют, Большая Пятерка (Big Five). Еще в 60-х годах рядом проведенных исследований, базирующихся на ранних работах Олпорта, Айзенка и Кеттелла, были обнаружены пять базовых личностных факторов [66]. Последующие исследования привели ученых к убеждению, что любая структурная модель личностных черт может быть сведена к пяти факторам, которые были названы Большой Пятеркой. В

наиболее законченном виде пятифакторная модель личности была оформлена в работах американских психологов Пола Косты и Роберта МакКрея [54], [64]. Авторы выделяют пять базовых параметров, с помощью которых может быть описана личность: *нейротизм, экстраверсия, открытость опыту, сотрудничество и добросовестность.*

Приведем краткое описание основных параметров личности.

Нейротизм. Данный параметр является наиболее всеобъемлющим по сравнению с другими. Он отражает способность индивида приспосабливаться к жизни, противостоять стрессовым ситуациям. Общая тенденция испытывать негативные аффективные состояния, такие как страх, грусть, раздражение и другие, составляет суть данной шкалы.

Экстраверсия. Экстраверсия, конечно, предполагает прежде всего общительность, но это только одна из черт, которую включает в себя данный фактор. Кроме того, экстраверты любят людей, большие группы и собрания, они напористы, активны, оптимистичны. Интроверты более сложны в описании. Интроверсия должна рассматриваться не как противоположность экстраверсии, а скорее как ее отсутствие.

Открытость опыту. Элементами открытости опыту являются такие личностные характеристики, как активное воображение, эстетическая чувствительность, внимание к чувствам других, гибкость ума, независимость в суждениях и оценках. Открытые индивиды проявляют интерес к внешнему и внутреннему миру, принимают и поддерживают новые идеи и нетрадиционные ценности.

Сотрудничество. Как и Экстраверсия, Сотрудничество является параметром межличностных тенденций. Сотрудничающий человек по своей природе является альтруистом. Он симпатизирует другим и стремится помочь им и верит, что и другие, в случае необходимости, не откажут ему в помощи. В отличие от него, соперничающий человек эгоцентричен, скептически относится к намерениям других и по духу является скорее конкурирующим, чем сотрудничающим.

Добросовестность. Это аспект личности, который был назван однажды «характером». Высокие баллы отражают скрупу-

лезность, пунктуальность, последовательность и надежность. Низкие баллы не обязательно означают отсутствие моральных принципов, просто такие люди менее последовательны в их применении. Кроме того, они более вялые и апатичные в достижении своих целей.

Представленная пятифакторная модель является чрезвычайно привлекательной для исследователей, что порождает соблазн изучать только указанные пять факторов, вместо сотен характеристик, исследуемых в настоящее время. Однако она тоже не свободна от недостатков [24].

1. До сих пор продолжается дискуссия относительно природы пятифакторной модели. Например, имеется точка зрения, согласно которой указанные пять параметров являются производными нашего языка. Таким образом, хотя личность в реальности может быть гораздо более сложной, наша способность описать личностные черты ограничивается прилагательными, имеющимися в нашем языке. Возможно и то, что наши когнитивные способности организовывать информацию о себе, ограничены использованием лишь пяти этих категорий. И хотя все личностные черты можно объединить в эти пять факторов, они все-таки не смогут отразить сложность, многообразие человеческой личности [24].

2. Полезность изучения глобальных личностных параметров не вызывает сомнения. Например, общительность и смелость могут быть связаны с более широким личностным фактором – экстраверсией. Однако если мы хотим понять поведение людей в конкретных социальных ситуациях, то лучше рассматривать их с точки зрения конкретных личностных характеристик, а не ориентироваться на глобальные факторы.

3. Наконец, существуют значительные расхождения мнений по поводу валидности пятифакторной модели личности. Например, в ряде факторных исследований были обнаружены паттерны характеристик, не укладывающиеся в ее рамки.

Указанные ограничения оставляют вопрос об оптимуме личностных характеристик открытым, однако пятифакторная модель личности является устойчивой и служит хорошей базой для создания тестов.

Авторами пятифакторной модели был разработан многофакторный личностный опросник для измерения личностных факторов Большой пятерки, который в окончательном варианте называется «Переработанный личностный опросник NEO» («NEO Personality Inventory Revised») (NEO PI-R) [54]. Он состоит из 240 вопросов, объединенных в 5 базовых шкал (domains). Каждая базовая шкала включает 6 подшкал (facets), описывающих более конкретные черты (Таблица 1).

Таблица 1

Шкалы и подшкалы, измеряемые опросником NEO PI – R

N : Нейротизм	N1: Тревожность N2: Враждебность N3: Депрессия	N4: Рефлексия N5: Импульсивность N6: Ранимость
E: Экстраверсия	E1:Сердечность E2: Общительность E3: Настойчивость	E4: Активность E5: Поиск возбуждения E6: Позитивные эмоции
O: Открытость опыту	O1: Фантазия O2: Эстетика O3: Чувства	O4: Действия O5: Идеи O6: Ценности
A: Сотрудничество	A1: Доверие A2: Честность A3: Альтруизм	A4: Уступчивость A5: Скромность A6: Чуткость
C: Добросовестность	C1: Компетентность C2: Организованность C3: Послушность долгу	C4: Стремление к достижениям C5: Самодисциплина C6: Обдумывание поступков

Авторы опросника считают его хорошим инструментом для диагностики личности, что подтверждается рядом исследований.

Во-первых, результаты методики NEO PI-R хорошо коррелируют с оценками по другим тестам, особенно с теми, которые построены на базе факторного анализа (опросники Г. Айзенка и Р. Кеттелла), что доказывает общность старых факторных моделей и Большой пятерки [25]. Аналогичные данные получены и в отечественных исследованиях [47].

Во-вторых, существуют доказательства корреляции данных NEO PI-R с результатами оценки личностных черт, полученных с

помощью других методов, например Q– сортировки или психосемантических методов [24], [47].

В-третьих, созданные адаптированные версии опросника NEO PI-R в разных языковых культурах (немецкой, итальянской, китайской, португальской, испанской, японской, польской, русской[4] и т.п.) и проведенные на их базе кросс-культурные исследования подтверждают общность структуры личности у представителей разных культур [55], [58].

Кроме того, в ряде исследований, в том числе и отечественных, обнаружено значительное совпадение самооценок испытуемых с теми оценками, которые давали испытуемым их близкие родственники (родные братья и сестры) или супруги [24], [63].

Все это позволяет предположить, что Большая Пятерка, измеряемая опросником NEO PI-R, необходима и достаточна для описания базовых параметров личности и «ни одна другая система не является столь экономной» [цит. по 24, р. 293].

Вторым подходом в русле теории личностных диспозиций является ***типологический подход***. Типы объединяют группы похожих людей и составляют набор объяснительных понятий, где в качестве названия выступает название соответствующего типа, а содержание раскрывается описанием типичного представителя. В отличие от предыдущего подхода, где единицей анализа является черта как некое психическое образование, в теории типов такой единицей анализа служит субъект как типичный представитель группы похожих людей (типа).

К числу наиболее известных типологий, ставших уже хрестоматийными, относятся классификации Э. Кречмера и В. Шелдона, в которых сделана попытка связать характер реакций человека с его физической конституцией. Более психологичной является типология, предложенная известным швейцарским медиком и психологом Карлом Густавом Юнгом (1875 – 1961). Он исходил из постулата о наличии у каждого человека преимуществен-

---

[4] Русскоязычная версия опросника NEO PI-R разработана авторами данного учебного пособия с разрешения издательской компании PAR и согласия авторов опросника совместно с американскими коллегами (см. [23], [63]).

но одного из двух способов связи человека с окружающим миром: экстравертированного и интровертированного. В случае экстраверсии человек направляет свою энергию на предметы окружающего мира. Такие люди активны, заинтересованы в контактах с окружающими, обладают некоей долей авантюризма. В случае интроверсии базовая ориентация направлена вовнутрь, на себя. Такой тип личности – осторожный, рефлексивный, неуверенный.

Предложенная Юнгом типология вскоре была им дополнена и конкретизирована введением четырех типов умственных операций, которые присутствуют в каждом человеке в разной степени. Два типа операций являются рациональными: ощущение и мышление, а остальных два – иррациональными. К ним относятся чувства и интуиция. Указанные два типа базисных ориентаций и 4 типа функций создают 8 различных типов личности. Хотя Юнг и признавал, что не все люди, входящие в ту или иную категорию, абсолютно похожи, но, тем не менее, не отрицал мысли о том, что двухчетвертная конструкция отчетливо присутствует в человеческой психике [48].

На основании юнгианской типологии двумя американскими психологами Катарин Бриггс и ее дочерью Илизабет Бриггс-Майерс в 1962 году был создан тест для определения личностного типа Myers-Briggs Type Indicator (MBTI), широко используемый в психодиагностике.

Другим примером типологического подхода может служить концепция профессиональных типов личности, предложенная американским психологом Джоном Холландом (1966). Автор считал, что люди, имеющие одинаковые предпочтения, установки, интересы, хобби, личностные особенности, склонны выбирать схожие профессиональные сферы. На основании этого им было предложено 6 типов личности с точки зрения их профессиональной направленности: реалистический, интеллектуальный, социальный, конвенциональный, предпринимательский и артистический. Каждый из этих типов заполняет собой соответствующую профессиональную среду, где занимает господствующее положение. Согласно Холланду, конкретные люди не принадлежат к строго определенному типу, а содержат в себе признаки всех ука-

занных типов в той или иной пропорции. Это позволяет описать любую личность в виде кодов, отражающих сочетание типов [62].

Предложенная автором типология была не только положена в основу большинства известных методик для диагностики профессиональных интересов (Опросник интересов Э. Стронга, Обозрение профессиональных интересов Ф. Кюдера), но и стала базисом для разработки собственной методики самоанализа профессиональных склонностей Self-Directed Search (SDS), которая с успехом применяется в практике профессионального консультирования как за рубежом, так и в нашей стране [3], [22], [31].

Типологический подход имеет ряд привлекательных сторон по сравнению с теорией черт. Прежде всего, он является эмоционально привлекательным, поскольку позволяет определить личностные особенности конкретного человека, отнеся его к тому или иному типу. Эта особенность делает данный подход более понятным для большинства людей.

Однако предложенные типологии не способны охватить все многообразие и сложность человеческого поведения и втиснуть богатство индивидуальностей в рамки какого-либо типа.

Подводя краткий итог анализу концепции личностных диспозиций, следует отметить ее сильные и слабые стороны.

К достоинствам концепции личностных диспозиций можно отнести следующие:

1. Теория личностных диспозиций, основываясь на принципе «надситуативности», дает возможность предсказывать поведение личности в той или иной ситуации.

2. Основная цель исследователя с точки зрения данной концепции заключается в описании личности, что позволяет не только показать сам факт отличия одного человека от другого по какой-либо личностной особенности, но дает возможность выразить эти различия в количественной форме.

Основным недостатком концепции считают невозможность объяснить поведение личности. Теория черт в большей степени описывает поведение человека, нежели объясняет его. Эта проблема беспокоит не только критиков теории, но и ее сторонников. В частности Г. Айзенк считал, «что без хорошей теории мы можем попасть в порочный круг: мы будем использовать понятие

черты для объяснения поведения, которое само первоначально служит основанием для того, чтобы выделить эту черту» [24, с. 323]. В самом деле, нанесение физического удара Васей Пете может служить проявлением такой личностной черты, как агрессивность. Но когда мы задаемся вопросом «почему Вася агрессивный?», то неизбежно приходим к ответу: «потому что он ударил Петю». Это свидетельствует о том, что теория черт не является объяснительной концепцией.

Кроме того, сомнению подвергается и мощность основного метода построения различных личностных моделей – факторного анализа. Это проявляется в неутихающих спорах о необходимом и достаточном количестве базовых черт. Хотя и достигнут некоторый консенсус по Большой Пятерке как наиболее оптимальной модели, однако степень соответствия между результатами разных исследований не идеальная. Некоторые авторы по-прежнему утверждают о возможности описать личность либо меньшим, либо большим количеством факторов [24].

Несмотря на указанные недостатки, теория личностных диспозиций в силу способности измерения личностных особенностей и возможности предсказания поведения и эффективности деятельности субъекта является основной теоретической и методологической базой традиционной психодиагностики, и подавляющее большинство всех известных тестов и конструирование новых методик проводится на ее основе.

3.2. Ситуационизм

Ситуационистский подход представляет собой альтернативную позицию по вопросу о том, что является ведущей детерминантой в связке «человек – ситуация» и соответственно предметом диагностики. Критики теории черт утверждают, что поведение гораздо сильнее варьирует от ситуации к ситуации, чем это предполагается концепцией личностных диспозиций. С их точки зрения наиболее эффективным является изучение не общей предрасположенности личности, а роли ситуации или вознаграждения со стороны среды. В литературных обзорах Уолтера Мишеля (W. Michel) и Дэвида Питерсона (D. Peterson) было показано, что

предсказуемость индивидуальных реакций в конкретных ситуациях очень низкая. В частности, коэффициенты корреляции между такой личностной чертой, как импульсивность, и ее различными поведенческими проявлениями находились в диапазоне 0,1– 0,2, а иногда и ниже [40]. Представители радикального ситуационизма считают, что предметом психодиагностики должен быть анализ поведения индивидов в зависимости от конкретных ситуаций, классификация ситуаций, определение степени влияния ситуации на поведенческие реакции.

Начавшаяся в конце 60-х годов XX в. дискуссия относительно преимущественной детерминации поведения человека личностными чертами и особенностями внешней ситуации дала начало печальному периоду в развитии психодиагностики, выразившемуся в борьбе двух альтернативных направлений: теории черт и ситуационизма. Результатом этой борьбы явилось появление третьего направления в изучении взаимодействия личности и ситуации – интеракционизма.

3.3. Интеракционизм

Интеракционизм предполагает рассмотрение не отдельно личностных черт и не отдельно характера ситуации как детерминант поведения, а их взаимодействия. В наибольшей степени эта позиция представлена в работах шведского психолога Д. Магнуссона. Магнуссон рассматривает поведение человека как процесс непрерывного взаимодействия между личностью и ситуацией. И хотя личностные особенности остаются существенными детерминантами поведения, однако большую роль играет психологическое значение ситуаций. В интеракционистской концепции происходит трансформация и основных постулатов радикального ситуационизма. Прежде всего, изменяется определение самого термина «ситуация». Понимание ситуации как совокупности элементов среды заменяется понятиями «субъективная ситуация», «значимая ситуация», «интерпретация ситуации» и т.п., тем самым подчеркивается перцептивный ее аспект, восприятие ситуации субъектом [5]. По мнению Д. Магнуссона, анализ ситуаций должен включать в себя анализ восприятия ситуации

субъектом, рассмотрение мотивационной стороны и анализ реакций на ситуацию.

Такое понимание ситуаций накладывает свой отпечаток и на их классификацию. В частности, в гипотезе У. Мишеля[5] о «сильных» и «слабых ситуациях» утверждается, что в зависимости от степени выраженности ситуации роль личностных и ситуационных переменных будет различна. В «сильных» ситуациях поведение будет меньше всего индивидуализировано, а в «слабых» – наоборот. Таким образом, в реакциях на сильные ситуации отмечаются незначительные вариации индивидуального поведения, что означает большую роль ситуационных факторов. Реакции на слабые ситуации более разнообразны, что говорит о большой значимости личностных особенностей.

Интеракционистский подход к личностно-ситуационному взаимодействию и пониманию ситуации как когнитивного конструкта отчетливо воплотился в теории личности известного американского психолога Джорджа Келли (1905 – 1996). Основным понятием его концепции является понятие конструкта, представляющего собой особое субъективное средство, созданное самим человеком, проверенное им на практике и помогающее ему воспринимать и понимать окружающий мир, прогнозировать и оценивать события. Личность человека представляет систему конструктов, с помощью которых он интерпретирует окружающую действительность и принимает решения о поведении в различных ситуациях. Индивиды отличаются друг от друга по содержанию конструктов, их количеству и степени организации в систему. Поэтому для того чтобы понять человека, надо понять систему его конструктов, т.е. узнать о тех событиях, которые включены в конструкты, о способах функционирования конструктов и о способах их организации в систему.

Представление о личности определило и основной методологический принцип их изучения, который был сформулирован Дж. Келли следующим образом: «если вы не знаете, что происходит в голове человека, спросите его, он может сказать вам» [11, с. 330].

[5] В последних своих работах У. Мишель отошел от концепции радикального ситуационизма и пересмотрел свои взгляды в пользу интеракционизма [5].

На основании этого автором был предложен собственный метод диагностики личности – Тест репертуарных решеток (Repertory Grid Technique), широко применяемый в зарубежной и отечественной психологии [4], [10], [13], [25], [30], [44], [47].

В отличие от теории черт, делающей упор на ситуационно неизменных и стабильных особенностях личности, подход Дж. Келли акцентирует внимание на человеке как активном агенте, постоянно находящемся в процессе изменения. Келли, как и остальные представители интеракционизма, считал, что человек ведет себя различным образом в разных ситуациях потому, что каждая ситуация конструируется и интерпретируется им по-разному. Он является сторонником идеографического подхода, поскольку делает акцент на уникальных аспектах системы конструктов каждого человека.

Несмотря на достоинства интеракционизма как объяснительного принципа взаимодействия личности и ситуации, на его основе создан лишь ограниченный класс методик, связанный с областью психосемантики.

При рассмотрении двух основных подходов к предмету психодиагностики личности в лице теории личностных диспозиций и интеракционизма (радикальный ситуационизм исключает личность как объект анализа) возникает вопрос о том, а единственно ли это возможные подходы к диагностике личностных особенностей.

Своеобразным ответом на этот вопрос послужила схема, предложенная российскими учеными А.Г. Шмелевым и В.И. Похилько [21] (Рис. 1).

Основная идея данной схемы заключается не в противопоставлении альтернативных подходов, а в создании некоей шкалы психодиагностических задач, расположенных по двум векторам «Уникальные – Общие свойства» и «Стабильные – Изменчивые свойства». Задачи I и III отражают соответственно номотетический и идеографический подходы в психодиагностике, воплощенные в теории личностных диспозиций и интеракционизме. Задачи II и IV занимают промежуточное положение между альтернативными подходами, и, хотя не имеют пока психологического инструментария для своего решения, но это не означает не-

возможности его создания, что очерчивает перспективы дальнейшего совершенствования методологии психодиагностики.

	Стабильные свойства		
Уникальные свойст-ва	II	I	Общие свойства
	III	IV	
	Изменчивые свойства		

Рис. 1. Схема психодиагностических задач

Традиционная психодиагностика, как было указано выше, основывается на теории личностных диспозиций, которая определяет всю процедуру создания и проверки теста. Данная процедура будет изложена в последующих разделах учебного пособия.

Вопросы для самопроверки

1. В чем основные достоинства и недостатки теории черт?
2. Представителем какого подхода в науке был Г. Олпорт и в чем сущность данного подхода?
3. Какой метод был положен в основу созданий концепции личности Р. Кеттелла и Г. Айзенка?
4. В чем сходство и различие между типологическим подходом и теорией черт?
5. В чем сущность интеракционизма?

Глава 4. **Определение тестов и их классификация**

Происхождение термина «тест» имеет давнюю историю. Считается, что слово «тест» берет свое начало от старофранцузского языка и является синонимом слова «testa» (чашка). Этим словом обозначались небольшие сосуды, использовавшиеся древними алхимиками для проведения химических опытов.

В русском языке слово тест имеет два значения. Во-первых, так называлась клятва, которую давали члены английской религиозной общины, демонстрируя тем самым, что они не являются скрытыми католиками. Во-вторых, словом «тест» называли плавильные сосуды, использовавшиеся для выделения олова из золота и серебра [4].

Таким образом, история происхождения термина «тест» подчеркивает главную его сущность: тест является измерительной пробой или испытанием. Именно в таком аспекте он применяется в различных областях науки и техники: медицине, биологии, инженерном деле.

В психологии существует более конкретное понимание теста как метода. А. Анастази под тестом понимает «объективное и стандартизированное измерение образцов (или проб) поведения» [2, с. 18], подчеркивая две его главные характеристики: объективность и стандартизированность. П. Клайн выделяет еще одну важную особенность теста, определяя его как «стандартизированное и часто ограниченное во времени испытание, предназначенное для установления количественных и качественных индивидуально-психологических особенностей» [12, с. 15].

Тест как самостоятельный метод исследования психических характеристик человека отличается от остальных методов психологии, прежде всего эксперимента.

1. Тест предназначен для диагностики индивидуальных различий и имеет в основном практическую направленность. Этим

он отличается от эксперимента, который является, прежде всего, исследовательской процедурой и служит для выявления общих закономерностей функционирования психики.

2. Тест, в отличие от эксперимента, представляет собой более кратковременную и более дешевую и экономически выгодную процедуру.

3. Вся процедура проведения теста, его содержание и способы обработки и интерпретации жестко стандартизированы, в то время как эксперимент предполагает внесение различного рода изменений со стороны исследователя.

4. Результаты тестирования предполагают сравнение с нормативными данными или критерием, что является обязательным условием конструирования любого теста.

Тесты можно классифицировать по-разному: по форме, цели, содержанию и другим характеристикам.

По *цели* тесты делятся на два класса. К первому классу относятся тесты, целью которых является измерение максимальных возможностей испытуемого. К этому классу относятся тесты интеллекта, способностей. Во вторую группу входят тесты, цель которых – определение типичного исполнения испытуемого, т.е. его поведения и деятельности в какой-либо ситуации или широком классе ситуаций. К этой категории относятся личностные тесты, тесты навыков, интересов [56].

По *материалу оперирования* тесты делятся на три группы: бланковые, предметные и аппаратурные. К бланковым тестам – тестам «карандаша и бумаги» – относятся личностные опросники, некоторые тесты интеллекта и способностей (тест структуры интеллекта (ТСИ) Р. Амтхауэра, тест «Домино», тест технических способностей Беннета). Предметные тесты предполагают работу испытуемого с какими-либо предметами (кубики Косса, тест «Сложение фигур» из батареи Векслера и т.п.). Аппаратурные тесты используются в основном для диагностики восприятия, памяти, мышления.

В соответствии с *количеством обследуемых* выделяют индивидуальные и групповые тесты. К первой группе относятся многофакторные личностные опросники, батареи для диагностики интеллекта Д. Векслера и Р. Амтхауэра, проективные тесты. Ин-

дивидуальные тесты требуют большего времени проведения и более высокой квалификации экспериментаторов, что обеспечивает высокую достоверность результатов. Они применяются в клинической практике, индивидуальном консультировании и на начальных этапах многих психологических исследований. К групповым тестам относятся стандартизированные личностные опросники, тесты интересов, тесты специальных способностей. Процедура их проведения более упрощена, но имеет и более жесткую стандартизацию. Квалификация экспериментатора может быть невысокой.

По характеру умственных действий тесты делятся на вербальные и невербальные. Вербальные тесты предполагают умение работать с вербальными символами, применять языковые правила. В эту группу относятся словесно-логические тесты, личностные опросники и т.п. Содержанием невербальных тестов является выявление умений оперировать образами или конкретными предметами (блоками, карточками, деталями). Примером таких тестов могут служить субтесты из теста структуры интеллекта Векслера: кубики Косса, «Сложение фигур» и т.п. [7].

В зависимости от *содержания измеряемой психической реальности* выделяют самые разнообразные классификации тестов.

Одной из первых классификаций методов сбора информации о личности была классификация, предложенная Р. Кеттеллом [19]. Он считал, что информация о личности может быть получена из разных источников и данные можно обозначить как «L»-, «Q»– и «T»-данные.

«L»-данные получили свое название от английских слов «life record data» и представляют собой регистрацию реального поведения человека в повседневной жизни. Обычно «L»-данные получают путем оценок экспертов, наблюдающих поведение испытуемых в течение некоторого периода времени в различных ситуациях. *«Q»-данные* (от Questionnaire data) предполагают изучение личности на основе самооценок и отображаются в различного рода опросниках. Благодаря простоте инструментария и легкости получения информации данная группа методов занимает центральное место в изучении личности. В эту группу входят

все известные личностные опросники, шкалы самооценок, шкалы мнений и установок и т.п.

«Т»-данные получили свое название от «objective test data» и представляют собой результаты объективных тестов с контролируемой экспериментальной ситуацией. Их получают путем объективного измерения поведения человека (вербального и невербального, индивидуального или социального и т.п.) без обращения к мнению экспертов. Общей особенностью всех объективных тестов является создание экспериментальной микроситуации, способствующей проявлению измеряемого психического свойства. В эту группу автор включает тесты способностей, тесты достижений, тесты на восприятие, ситуационные тесты, проективные тесты и игры. Кроме того, сюда также включены физиологические тесты, в которых регистрируются физиологические показатели (КГР, ЭКГ, ЭЭГ, тремор и т.п.) и физические тесты, связанные с антропометрическими показателями человека: ростом, весом, длиной конечностей, объемом груди и т.п.

Похожий подход можно найти и в отечественной психодиагностике. В частности, Л.Ф. Бурлачук и С.М. Морозов выделяют три основных подхода к диагностике психического, которые охватывают все множество существующих тестов: объективный подход, субъективный подход и проективный подход.

Объективный подход предполагает измерение результативности (успешности) и способа выполнения деятельности. В эту группу относятся тесты интеллекта, способностей и достижений, а также некоторые личностные тесты (ситуационные и тесты действий).

Субъективный подход основан на самооценке субъектом своих чувств, особенностей поведения в разных ситуациях. Он представлен многочисленными опросниками. Эти диагностические процедуры подразделяются на личностные опросники, опросники состояний и настроения, а также опросники мнений и опросники-анкеты. Последние две группы не относятся собственно к личностным опросникам, а используются в основном при социологических исследованиях.

Проективный подход предполагает диагностику личности на основе взаимодействия субъекта с нейтральным, безличным ма-

териалом, который в силу особенностей своей неопределенной структуры становится объектом проекции чувств, мотивов и особенностей личности. В эту группу входят все известные проективные методики [3]. Существует несколько достаточно популярных классификаций проективных тестов, среди которых наиболее удобной, по мнению авторов, является классификация, предложенная Саулом Розенцвейгом в 1964 году. Он выделял три группы проективных методик: моторно-экспрессивные, перцептивно-структурные и апперцептивно-динамические [3].

Другие авторы предлагают более дифференцированные классификации тестов. А. Анастази выделяет следующие группы тестов: 1. Интеллектуальные тесты. 2. Тесты способностей. 3. Тесты достижений. 4. Личностные тесты. 5. Прочие методики диагностики психических свойств [2].

Интеллектуальные тесты подразделяются на индивидуальные тесты, групповые тесты и тесты для специфических популяций, к которым относятся тесты для больных, страдающих нарушением анализаторных систем (зрения, слуха) и опорно-двигательного аппарата, а также тесты для испытуемых других культур или неграмотных.

Тесты способностей включают две группы: комплексные батареи диагностики способностей и тесты отдельных специальных способностей (музыкальных, технических, художественных и т.п.).

Группа личностных тестов состоит из опросников личностных черт, тестов интересов и установок и проективных методик.

Наконец к последней группе тестов относят тесты для диагностики личностных особенностей, которые не связаны с самооценкой личности. Это тесты действия, ситуационные тесты, а также тесты представлений субъекта о себе.

Обзор приведенных выше классификаций тестов показывает, что любой тест является некоей точкой многомерного пространства оснований и может иметь различные характеристики [7].

Тесты не создаются одинаково, и критериями «хорошего» теста не являются случайные факторы. Создание теста – это результат процесса мышления, с одной стороны, и соблюдения определенных правил конструирования – с другой. Это значит, что

процесс конструирования теста включает в себя как творческие компоненты, особенно на первых этапах, так и рутинные, поддающиеся алгоритмическому описанию.

Каковы же основные признаки «хорошего» теста?

Психологический тест может быть охарактеризован как эффективный инструмент при условии, что он удовлетворяет определенным требованиям. Основные из них следующие [12]:

1. Использование психометрической шкалы не ниже интервальной.

2. Высокая надежность теста.

3. Высокий уровень валидности теста.

4. Хорошая различительная способность каждого задания теста, т.е. дискриминативность.

5. Наличие нормативных данных, или возможность описания результатов теста с высокой точностью различными моделями, или наличие критериев установленных экспертами.

Для достижения этих требований и создания действительно высокоэффективного теста необходимо осуществление ряда процедур при его конструировании. Традиционно выделяют три этапа создания любого теста.

I. Разработка первичной формы теста. Данный этап включает в себя следующие процедуры:

1. Выбор предмета диагностики.

2. Анализ и теоретическая разработка предмета диагностики, которая заключается в создании теоретического и операционального определения предмета диагностики.

3. Выбор области валидности применяемого теста.

4. Выбор шкалы ответов.

5. Выбор типов заданий в соответствии с характером ответа.

6. Составление и формулирование заданий теста и определение порядка размещения их в тесте.

7. Проведение созданной формы на специально подобранной выборке.

II. Психометрическая проверка теста. Данный этап включает в себя следующие процедуры:

1. Анализ пунктов теста. Смысл этой процедуры состоит в том, чтобы определить, насколько каждое задание теста соответствует общему диагностическому содержанию методики.

2. Проверка основных видов надежности теста. Задачей данной процедуры является определение того, насколько результаты тестирования подвержены влиянию различных побочных факторов.

3. Оценка основных видов валидности или обоснованности теста. Задачей данной процедуры является определение того, насколько хорошо создаваемая методика выполняет свои диагностические задачи, т.е. диагностирует тот конструкт, который лежит в ее основе.

III. Расчет нормативных показателей.

Приведенная выше процедура разработки теста предполагает выполнение целого ряда мероприятий. В силу ограниченности объема нашего пособия далее мы остановимся только на тех этапах разработки теста, которые недостаточно полно отражены в отечественной психодиагностической литературе или вызывают наибольшие трудности при понимании материала.

Вопросы для самопроверки

1. Что такое тест?
2. Назовите несколько группировок тестов (3 - 4 классификации).
3. В чем отличие теста от экспериментального метода?
4. Перечислите основные признаки «хорошего» теста.

Глава 5. **Этапы психометрической проверки теста**

5.1. Основные статистические понятия

Прежде чем перейти к описанию основных этапов разработки и психометрической проверки психологического теста, кратко остановимся на описании основных статистических понятий, лежащих в их основе.

Закон нормального распределения. Одним из основных понятий математической статистики, на которой основан закон нормального распределении, является понятие случайной величины. Случайной называется переменная величина, способная в одних и тех же условиях принимать различные числовые значения. И в этом смысле количественные значения, отражающие изменения любого психического свойства, также можно считать случайными величинами. Случайные величины делятся на дискретные и непрерывные. Случайная величина называется дискретной, если она может принимать значения, выражающиеся только целыми числами. Если случайная величина способна принимать любые числовые значения, она называется непрерывной.

Естественно, что для каждого отдельного значения случайной величины можно рассчитать количество случаев, для которых и было зафиксировано данное значение. Если для каждого значения случайной величины мы рассчитаем количество случаев и представим это графически, то мы получим кривую распределения случайной величины. В том случае, если кривая распределения принимает вид симметричной колоколообразной кривой, мы можем говорить о том, что распределение нашей случайной величины имеет нормальный характер (рис. 2).



Рис. 2. Кривая нормального распределения

Итак, в том случае, если полученные эмпирические данные подчиняются закону нормального распределения, то график такого распределения будет представлять собой симметричную колоколообразную кривую. Кривая нормального распределения обозначает, что число случаев максимально в середине распределения и постепенно спадает к ее краям. Кривая симметрична и имеет единственный пик в центре. Большинство показателей – от роста и веса до способностей и параметров личности – приближаются к кривой нормального распределения. Этот тип кривой обладает важными математическими свойствами, и на ней основаны многие виды математического анализа. Для наших целей важны лишь некоторые из них.

Оценка нормальности распределения. Для статистического анализа результатов конкретных эмпирических исследований важно знать, подчиняется ли полученное распределение нормальному закону, поскольку в том случае, если полученное распределение не будет подчиняться нормальному закону, то никакие выводы о тех или иных закономерностях функционирования изучаемой переменной нельзя будет считать правомерными. Необходимо отметить, что отклонения эмпирических рядов распределения от нормальной кривой происходят довольно часто. Эти отклонения могут быть различными, обнаруживая в одних случаях асимметрию, в других эксцесс, а иногда и то и другое одновременно (методы расчета асимметрии и эксцесса описаны в Главе 6).

Меры центральной тенденции. В самом общем смысле меры центральной тенденции указывают единственный наиболее типичный результат той или иной переменной для всей группы испытуемых или, другими словами, для всех наблюдений. Самой известной из мер центральной тенденции является *среднее* (среднеарифметическое значение), обозначаемое либо М, либо x . Другой мерой центральной тенденции является *мода*, или наиболее часто встречающийся результат. Третья мера центральной тенденции – это *медиана.* Медиана – это средний показатель, относительно которого ряд распределения делится на две половины: в обе стороны от медианы располагается одинаковое число членов ряда. Для небольших выборок определить медиану довольно просто. Для этого всю совокупность наблюдений ранжируют по возрастающим значениям признака. Если число членов ряда нечетное, например: 1 2 3 4 5 6 7 8 9, то центральный вариант и будет его медианой. В данном случае: 5. Если число членов ряда четное, например 1 2 3 4 5 6 7 8 9 10, то медиана будет определяться по полусумме двух соседних вариантов, расположенных в центре ряда. В данном случае: 5,5. При идеально нормальном распределении медиана делит распределение точно пополам. Отметим, что меры центральной тенденции или средние величины не дают нам полной информации о варьирующих объектах, которыми в сущности являются и данные любых эмпирических психологических исследований. Поэтому наряду со средними величинами для более полной характеристики вариационного ряда должны вычисляться и показатели вариации или меры разброса данных.

Меры разброса данных. Наиболее наглядным и понятным способом представления разброса данных является *размах распределения*, т.е. разность между самым высоким и самым низким результатом. С одной стороны, мера размаха распределения понятна и проста, в чем заключается его положительная сторона. Однако, с другой стороны, данный показатель может сильно меняться при повторных выборках одной и той же генеральной совокупности, т.е. эта мера крайне неточна и неустойчива. Более точный результат измерения разброса основан на учете разности между каждым индивидуальным и средним значением по группе.

Наиболее часто употребляемой мерой центральной тенденции является *стандартное отклонение*, обозначаемое буквой σ (сигма). Стандартное отклонение вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum (x^i - \overline{x})^2}{n}},$$

где x^i – каждый индивидуальный показатель по группе,

$\overline{x}$ – среднее значение по группе,

n – общий объем выборки.

Чтобы понять, какую роль в законе нормального распределения играют меры центральной тенденции (в частности среднее значение M) и меры разброса данных (в частности стандартное отклонение σ), представим, что мы получили абсолютно нормальное распределение. В этом случае среднее значение будет совпадать с медианой. Если в разные стороны от среднего значения мы отложим значение σ, то в этом диапазоне в соответствии с законом находится 68% площади под кривой, а значит, 68% случаев наблюдаемой переменной. Если затем отложить в разные стороны еще по одному значению σ, то в этих диапазонах будет находиться по 14% случаев, а если отложить еще по одной σ, то в этих диапазонах будет находиться по 2% случаев (Рис. 3).

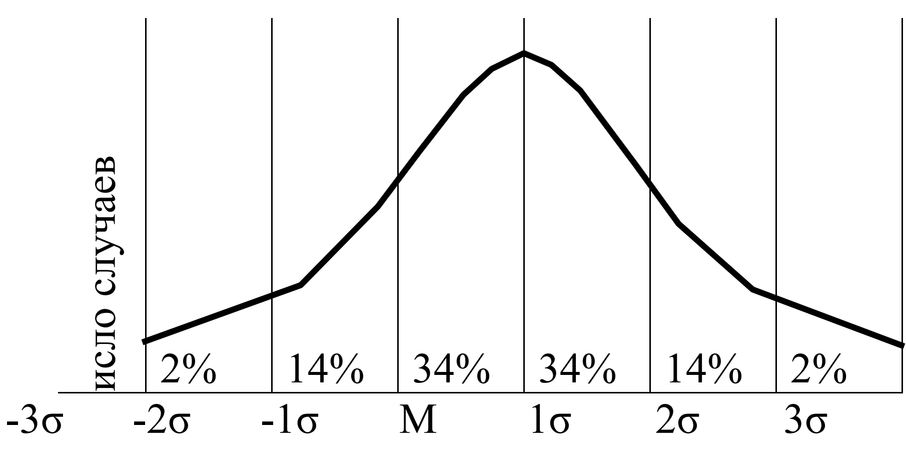


Рис. 3. Процентное распределение случаев на нормальной кривой

Все это приводит нас к еще одному положению закона нормального распределения – правилу трех сигм, которое заключа-

ется в том, что нормальное распределение должно строго укладываться в пределах трех сигм в одну и в другую сторону от среднего значения.

Уровень значимости. Показатель, используемый для проверки статистических гипотез. Уровень значимости отражает степень вероятности ошибочного вывода относительно статистической гипотезы оби признака, проверяемой на основе выборочных данных. Проверка статистических гипотез всегда связана с вероятностью неправильного, ошибочного суждения в силу действия каких-либо случайных непредвиденных факторов. Уровень значимости показывает значение этой вероятности, при котором действиями этих случайных, непредвиденных факторов можно пренебречь. Уровень значимости, как правило, обозначается буквой Р и выражается в процентах и долях вероятности ошибок. Уровень значимости и есть величина, оценивающая этот риск. В психологических исследованиях за достаточный уровень значимости обычно принимается Р=0,05, что означает, что в 5% случаев повторных выборок вывод, принимающий или отвергающий проверяемую гипотезу, может оказаться неправильным. Более «строгим» является уровень значимости 0.01, который означает, что лишь в 1% случаев повторных выборок вывод, принимающий или отвергающий проверяемую гипотезу, может оказаться неправильным.

Понятие корреляции. Для точного определения наличия или отсутствия связи между переменными используется понятие корреляции (от латинского Correlatio – соотношение). Корреляция – одна из наиболее часто употребляемых статистических процедур, применяемая при создании психологических тестов. С помощью данной процедуры определяется степень соответствия, связи между двумя переменными. Математическим выражением корреляции является коэффициент корреляции, обозначаемой буквой r. Коэффициент корреляции может находиться в пределах от –1 (когда между переменными обнаруживается полная обратная связь), до 1 (когда между переменными обнаруживается полная прямая связь). Коэффициент корреляции равный нулю говорит об отсутствии какой-либо связи между переменными.

Итак, значение коэффициента корреляции может колебаться в пределах от –1 до 1. Представим, что в нашем исследовании мы получили коэффициент корреляции между двумя переменными равный 0,20. Каким образом мы можем определить: существует ли связь между этими переменными, т.е. определить значимость данного коэффициента корреляции. Оценка значимости коэффициента корреляции производится с помощью специальных таблиц в зависимости от размера нашей выборки и уровня значимости (см. Приложение 1).

Интерпретация коэффициента корреляции. Наиболее распространенной ошибкой интерпретации любого коэффициента корреляции является попытка определить через него причинно-следственные связи между изучаемыми переменными. Отметим, что наличие корреляции двух переменных ни в коем случае не означает того, что между ними существует причинно-следственная связь. При этом, однако, необходимо учитывать природу тех явлений, которые являются переменными при подсчете корреляций. В некоторых случаях мы действительно имеем дело с измерением причины и следствия. Например Х – возраст человека, У – его общая осведомленность об окружающем мире, которая является одной из составляющих вербального интеллекта. Но все равно необходимо помнить, что даже в тех случаях, когда можно предположить существование причинно-следственной связи между двумя переменными, которые коррелируют между собой, сам по себе коэффициент корреляции ничего не говорит нам о том, вызывает ли Х появление У или наоборот. Кроме того, даже статистически выявленные связи в гуманитарных науках почти всегда слишком сложны, чтобы их объяснением могла служить какая-либо единственная причина. И чем сложнее и неопределенней природа изучаемой переменной, тем больше факторов влияет на ее связи с другими переменными.

Для вычисления коэффициента корреляции могут использоваться как параметрические, так и не параметрические критерии.

Параметрические критерии предполагают наличие нормальности распределения полученных данных и основаны на вычислении среднеарифметического значения и стандартного отклонения. Наиболее известным критерием является коэффициент ли-

нейной корреляции К. Пирсона, который используется для больших выборок и считается наиболее надежным и точным [18], [43].

Для малых выборок, где достижение нормального распределения данных весьма проблематично, используются непараметрические критерии, основанные на ранжировании конкретных результатов. Наиболее известным среди этой группы критериев является коэффициент ранговой корреляции Ч. Спирмена [18], [43].

Понятие достоверности различий. Определение достоверности различий – это статистическая процедура установления значимости различий между выборками по изучаемым переменным. Данная процедура позволяет получить подтверждение или опровержение гипотезы о том, что различия между выборками по какой-либо переменной являются статистически значимыми. Например, в ходе эмпирического исследования мы получили данные о том, что среднее значение по параметру экстраверсии для мужской выборки оказалось равным 26 баллам, а для женской – 29 баллам. Можем ли мы на основе этих данных сделать вывод о том, что женщины более экстравертированны, чем мужчины. На первый взгляд, мы можем сделать такой вывод, поскольку среднее значение для женской выборки оказалось на 3 балла большим, чем для мужской. Однако с точки зрения статистики такая разница в 3 балла может быть незначимой, а это, в свою очередь, будет говорить нам о том, что реальных различий между нашими выборками по параметру экстраверсии не существует. При оценке значимости различий используется целый ряд как параметрических, так и непараметрических критериев. Среди параметрических критериев это, в первую очередь, t-критерий Стьюдента и F-критерий Фишера. Среди непараметрических критериев это T-критерий Уайта, критерий χ^2 и другие [17], [42]. Оценка значимости различий также производится с помощью специальных таблиц в зависимости от размера выборки и выбранного уровня значимости. В приложении приведена таблица критических значений определения значимости различий для t-критерия Стьюдента (см. Приложение 2).

5.2. Выбор предмета диагностики

Перед психологом может стоять любая практическая задача, связанная с профессиональным отбором или профессиональным консультированием, уточнением клинического диагноза или отбором учащихся в профильные классы и т.п. В любом из этих случаев психологу необходимо разрабатывать тест. В зависимости от поставленной цели могут существовать два типа тестов: тесты***, не требующие создания гипотетического конструкта,*** и тесты***, для которых создание гипотетического конструкта является необходимым условием.***

В первом случае отсутствие гипотетического конструкта может обосновываться конкретной, чисто практической целью исследования. Например, для отбора радиотелеграфистов в рядах Вооруженных Сил в конце 70-х годов прошлого века использовалась батарея тестов, включавшая в себя известные методики для выявления особенностей психических процессов (внимания, памяти, способности оперирования наглядными представлениями и т.п.). В нее входили корректурные пробы, методика «Перепутанных линий», методики на запоминание числовых рядов, тест «Компасы» и некоторые другие.

Другим вариантом деятельности психолога, когда не требуется создание гипотетического конструкта, является использование теста с уже готовым гипотетическим конструктом. В этом случае мы имеем дело с адаптацией зарубежных тестов, которую хотелось бы рассмотреть несколько более подробно по ряду причин.

Во-первых, в отечественной практике сложилось, на наш взгляд, довольно распространенное мнение у большинства пользователей о том, что для адаптации зарубежной методики достаточно только осуществить ее перевод. Устранить грубые ошибки, касающиеся межкультурных различий, затем осуществлять психометрическую проверку (в лучшем случае) или сразу использовать тест для научных и практических целей без таковой.

Во-вторых, в современной отечественной психодиагностической литературе этому вопросу уделяется мало внимания.

Адаптация зарубежных методик – достаточно трудоемкий процесс, который предполагает, по мнению Ю.Л. Ханина, осуществление следующих этапов [3]: 1) подготовка первичного перевода с оригинала на родной язык; 2) экспертная содержательная оценка предварительного перевода с привлечением лингвистов и психологов, владеющих языком оригинала; 3) проверка эквивалентности экспериментального перевода шкалы и оригинала; 4) установление новой шкалы и сбор соответствующих норм на отечественной выборке.

Рассмотрим этапы адаптации зарубежных методик, взяв за основу указанную процедуру Ю.Л. Ханина и используя разработки, имеющиеся в мировой психодиагностике, и собственный опыт в адаптации зарубежных личностных опросников.

1. Подготовка первичного перевода теста с языка оригинала на другой язык. Данную процедуру должны осуществлять лица, свободно владеющие обоими языками (билингвы) и желательно имеющие психологическое образование. С одной стороны, данная процедура не вызывает дополнительных проблем в силу своей конкретности и понятности, но, с другой стороны, может создать трудности, одной из причин которых может быть «стоящая за каждым опросником психометрическая проработка оригинала, заставляющая считать его неприкосновенным» [3, с. 125]. В мировой практике различают два типа перевода: симметричный и асимметричный. *Симметричный перевод* предполагает сохранение как смысла утверждения, так и привычного, разговорного стиля. *Асимметричный перевод* нацелен на сохранение верности какому-либо языку, обычно языку оригинала, ориентируясь на точность смысла и дословность перевода. Разработчики, занимающиеся адаптацией, как правило, ориентируются на асимметричный перевод, желая сохранить верность оригиналу. Результатом такой деятельности является достаточно неуклюжая с лексической точки зрения конструкция, которая, по справедливому утверждению Л.Ф. Бурлачука, скорее направлена на понимание испытуемым смысла утверждений, чем на диагностику соответствующих личностных характеристик [3].

2. Экспертная содержательная оценка перевода предполагает адаптацию лексики и грамматики языка к возрасту и уровню об-

разования генеральной совокупности, для которой предназначен тест. На этом этапе устраняются наиболее грубые различия между оригиналом и переводом. При этом следует обращать внимание на культурные особенности и другие аспекты жизни того общества, в котором создавался оригинал теста.

3. Проверка эквивалентности перевода оригиналу. Данный этап предполагает соотнесение полученного перевода опросника с внесенными поправками лексического и грамматического плана с оригиналом. Конкретной процедурой такой проверки является обратный перевод с другого языка на язык оригинала. Данный перевод, по нашему мнению, должна осуществлять другая команда билингвов с психологической подготовкой, не участвующих на первом этапе. Именно такая процедура была использована нами при создании русскоязычной версии опросника NEO PI–R. На каждое утверждение нами была составлена табличка с указанием утверждения на английском языке, его первоначального перевода на русский и обратного перевода с русского на английский. При этом авторы русскоязычной версии не добивались дословного перевода, выражающегося в полном совпадении терминов. При обратном переводе с русского на английский возможным было употребление синонимов, если они точно отражали смысл утверждения и стиль его изложения. С точки зрения имеющейся мировой практики такой обратный перевод с одного языка на другой язык может осуществляться несколько раз до полного совпадения оригинала и перевода теста.

4. Установление новой шкалы и сбор соответствующих норм – на данном этапе, по мнению Ю.Л. Ханина, мы должны осуществить сбор новых нормативных данных на отечественной выборке. С нашей точки зрения, такой подход не совсем верен, и сбору нормативных данных должна предшествовать стандартная процедура психометрической проверки теста, поскольку после достижения эквивалентности перевода и оригинала и «преодолев лингвистические и социокультурные препятствия, в итоге получаем *новый опросник*» [3, с. 126]. Такая процедура включает анализ пунктов теста и проверку теста по основным психометрическим характеристикам: надежности, валидности. Кроме того, проведенный анализ пунктов теста и вычисление коэффициента

дискриминативности может дать дополнительную информацию о тех факторах, которые послужили причиной неспособности того или иного задания теста дифференцировать испытуемых на людей с высокими и низкими баллами по тесту. В этом аспекте анализ пунктов теста может служить процедурой для установления тех социокультурных различий между популяциями, которые не удалось выявить на этапе экспертной содержательной оценки. В частности, после установления эквивалентности перевода опросника NEO PI–R языку оригинала нами был проведен анализ пунктов теста, который выявил слабую различительную способность ряда утверждений оригинала. Как показал дальнейший анализ, причины низкой дискриминативности заданий теста заключались в имеющихся различиях в социально-экономической ситуации в США и России того периода[6], а также были обусловлены культурой общества. Основная причина низких коэффициентов дискриминативности заключалась в наличии социокультурных различий между русской и американской выборкой. Особенно чувствительной в этом отношении оказалась шкала А (Сотрудничество), где очень много утверждений касалось социальных и моральных установок людей. Например, неудовлетворительными оказались такие утверждения, как «Мы никогда не сможем сделать слишком много для бедных и престарелых» и «Я не испытываю сострадания к нищим». Все утверждения этой шкалы отражают установки, симпатии и интерес к другим людям, их потребностям и нуждам. Однако в силу сложившихся социально-экономических условий в России того периода, все, что связано с чуткостью и вниманием по отношению к другим людям, для большинства населения России имело второстепенное значение. На первом плане стояли проблемы улучшения своего материального положения, проблемы собственного выживания.

Вторая по значимости причина была связана с неудачным переводом отдельных утверждений (неточность передачи смысла утверждения) или несоответствием стилистическим правилам русского языка. Примером данной причины может служить ут-

[6] Начальные этапы адаптации методики NEO PI–R проводились в 1995 – 1996 году.

верждение «Я легко сопротивляюсь соблазнам». Данный перевод оказался не вполне удачным, поскольку в повседневном русском языке сочетание «легко сопротивляюсь» не употребляется. Другим примером неадекватного перевода можно считать случай, когда перевод неточно передавал смысл измеряемого конструкта, что было вызвано различиями в понимании некоторых понятий в русском и английском языке. Так, первичный перевод утверждения из шкалы N (Нейротизм) имел следующую формулировку: «Меня считают “горячим”, темпераментным человеком». Согласно измеряемому конструкту, высокий балл по этой шкале характеризует человека как враждебного, вспыльчивого, раздражительного, легко впадающего в состояние гнева. Однако в русском языке понятие «горячий», темпераментный человек не обязательно связывается с отрицательными эмоциями и может относиться к человеку активному, энергичному и жизнерадостному.

После проведения анализа пунктов теста и устранения «не сработавших» утверждений нами была осуществлена полная психометрическая проверка методики с вычислением ретестовой надежности и надежности по однородности, а также критериальной и конструктной валидности. Кроме того, поскольку мы имели дело с многофакторным личностным опросником, мы должны были осуществить проверку воспроизводимости структуры опросника, что и было сделано с помощью факторного анализа.

В заключение можно констатировать факт, что адаптация зарубежных тестов по своему объему и трудоемкости сопоставима с разработкой оригинального теста и отличие заключается в том, что вместе с методикой мы берем и ту теоретическую конструкцию, которая положена в ее основу.

В ряде случаев перед психологом-разработчиком задача формулируется достаточно неопределенно, когда ему предлагается создать методику для оценки основных характеристик личности. В этом случае ему нужно создать свой гипотетический конструкт, предполагающий формирование представления о структуре личностных черт, их содержании и основных проявлениях в поведении человека.

5.3. Теоретическая разработка предмета диагностики. Теоретическое и операциональное определение

Известно, что все психологические явления, которые зафиксированы такими понятиями, как «душа», «психика», «сознание», «мышление», «интеллект», «мотивация», «потребности» и т.д., являются гипотетическими конструкциями, которые вводятся для объяснения тех или иных аспектов поведения человека. Это значит, что существование этих явлений первоначально выдвигается как гипотеза для интерпретации и понимания изучаемых фактов, а затем в результате экспериментальных проверок и теоретических обоснований данные явления входят в багаж психологической науки. Некоторые гипотетические конструкции не выдерживают такого рода проверки, и поэтому со временем от них приходится отказываться. Другие же конструкты, наоборот, входят в арсенал используемых наукой понятий и постепенно приобретают статус объективно существующих явлений.

Идея гипотетической конструкции (гипотетической переменной, гипотетического конструкта) состоит в том, что прямо и непосредственно ни одно психическое явление мы наблюдать не можем. Реально наблюдать и фиксировать мы можем только поведение человека (речевое и моторное поведение, сомато-вегетативное поведение и т.п.). Для того чтобы объяснить те или иные особенности поведения, психологи вынуждены вводить предположения, гипотезы о существовании тех или иных психических процессов, свойств, состояний, наличие которых позволяет логично объяснить наблюдаемые поведенческие факты.

Например, наблюдая в самых разных ситуациях человека, который грубо, враждебно ведет себя по отношению к окружающим, мы говорим, что он обладает такой чертой личности, как *агрессивность.* В данном случае *агрессивность* – это гипотетический конструкт, который мы ввели для объяснения враждебного отношения к людям. Другой пример: когда мы видим, что после предъявления некоторых стимулов человек может через некото-

рое время их воспроизвести, то мы говорим о процессах запоминания и воспроизведения, а в целом о существовании такой психологической функции, как память. Здесь понятия *«процессы запоминания»* и *«процессы воспроизведения»* также являются гипотетическими конструкциями, которые еще на заре возникновения психологии как науки были введены для объяснения фактов сохранения и воспроизведения ранее предъявленных стимулов. С одной стороны, эти конструкты оказались удобными для объяснения ряда психических явлений, а с другой стороны, их существование было подтверждено результатами экспериментально-психологических обследований, также нейрофизиологическими данными о функциях нервных клеток. В настоящее время существование этих процессов рассматривается уже не как гипотеза, а как реально существующее явление.

Поэтому разработка нового психологического инструмента начинается с того, что на основе наблюдений за поведением людей, сбора тех поведенческих жизненных явлений, которые требуют объяснения и диагностики, автор создаваемого теста должен выдвинуть гипотезу о том, **что** лежит в основе тех или иных особенностей человеческого поведения, гипотезу о том, почему существуют различия в поведении разных людей и какая именно особенность личности определяет эти различия. На данном этапе дается также и название вновь вводимого гипотетического конструкта.

Необходимым этапом является также определение предмета диагностики вновь создаваемого теста. На данном этапе вводимому параметру необходимо дать *теоретическое* и *операциональное* определение [69].

Теоретическое определение диагностического конструкта – это определение предлагаемого конструкта в понятиях, категориях и связях уже существующих в данном разделе науки теорий, понятий и категорий. Теоретическое определение – это включение предлагаемого понятия в уже существующую систему научных понятий в данной области психологии. Дать теоретическое определение – это значит прочно связать предлагаемый гипотетический конструкт с уже существующими и доказавшими свое право на существование конструктами и понятиями.

Операциональное определение диагностического конструкта – это определение, даваемое в терминах действий, операций и поведенческих актов. Операциональное определение – это описание тех поведенческих проявлений, реакций и действий, чувств и оценок, посредством которых данная переменная реализуется. Операциональное определение выводит изучаемую переменную на уровень фиксируемых, наблюдаемых и измеряемых проявлений и фактов, что делает ее доступной изучению и анализу. Кроме того, операциональное определение является непосредственной основой для построения первичной формы создаваемого теста.

В практике мировой психодиагносики для операционализации гипотетического конструкта обычно используют следующие проявления:

1. Поведенческие реакции, действия, поступки. Например, гипотетический конструкт черты личности «общительность» может проявляться в количестве знакомых людей, в активности человека в ситуации общения, в степени его разговорчивости. Примером поведенческих проявлений может служить утверждение: *«Я обычно первым начинаю разговор с соседями по купе в поезде»*.

2. Мотивационные компоненты, проявляющиеся в предпочтениях, желаниях и побуждениях. Та же самая общительность может проявляться в желании человека бывать в компаниях людей, в предпочтениях людских собраний и т.д. Пример*: «Я предпочитаю проводить свой отпуск на шумных курортах, нежели в уединенной хижине в горах»*

3. Эмоциональные проявления, выражающиеся в получении положительных эмоций в ситуациях, где проявляется данная личностная черта или наоборот. Пример: *«Одиночество меня пугает»*.

4. Оценочные проявления, выражающиеся в самооценке своих личностных особенностей или в их оценке со стороны других людей. Пример: *«Я – общительный человек»* или *«Меня считают общительным человеком»*.

Указанные проявления могут стать основой для формулировки будущих заданий. При этом не воспрещается в качестве ис-

точника формулировки заданий теста использовать задачи и утверждения из тестов, сходных по своему характеру с разрабатываемым. В частности, анализ известных личностных опросников, таких как 16 PF Р. Кеттелла, опросник Г. Айзенка и NEO PI – R, показывает заимствование одних и тех же утверждений. Для клинических тестов можно использовать в качестве концептов заданий описание симптомов психических заболеваний, интересующих исследователя [7].

5.4. Анализ заданий теста

Первый этап психометрической проверки психодиагностической методики заключается в определении диагностической пригодности каждого из заданий методики. Работа по данному этапу начинается после того, как первичная форма теста будет разработана и проведена на специально подобранной выборке. Предполагается, что на данном этапе каждое задание психологического теста должно быть проверено по нескольким психометрическим показателям.

1. Коэффициент корреляции между ответами испытуемого на каждый отдельный пункт опросника и итоговым баллом по той шкале, в которую входит данный пункт, должен быть статистически значимым, поскольку в этом случае данный коэффициент корреляции отражает тот факт, что данное конкретное утверждение способно измерять тот же самый конструкт, что измеряет вся шкала в целом, в которую, естественно, входит данный пункт. В литературе, посвященной проблемам психометрики, данный психометрический параметр известен как «дискриминативность заданий теста», т.е. способность отдельных заданий теста дифференцировать испытуемых относительно «минимального» или «максимального» результата теста. Коэффициент дискриминативности, как и любой коэффициент корреляции, может принимать значение от –1 до +1.

Значимый положительный коэффициент дискриминативности свидетельствует о том, что данное задание направлено на измерение того же параметра личности, который измеряет вся шкала в целом, а, кроме того, это говорит о хорошей способности

данного задания для деления испытуемых. В свою очередь, значимое отрицательное значение данного коэффициента, с одной стороны, говорит о том, что данное задание направлено на измерение параметра личности, который противоположен тому, который измеряет вся шкала. Отметим, что значимое отрицательное значение данного коэффициента далеко не всегда говорит о том, что такое задание подлежит немедленному исключению из опросника, поскольку так называемые «обратные задания» должны иметь именно отрицательные и значимые коэффициенты дискриминативности. В том случае, если коэффициент дискриминативности не является значимым, это говорит о том, что данное задание не выполняет свою диагностическую задачу, поскольку измеряемое им свойство личности никак не связано с тем свойством личности, которое измеряет вся шкала в целом.

По мнению ряда авторов, коэффициент дискриминативности заданий теста является, по сути, показателем конструктной валидности отдельного пункта, поскольку определяется по отношению к внешнему критерию – суммарному результату по всему тесту.

2. Индекс трудности, смысл которого состоит в том, чтобы определить процентное соотношение испытуемых, ответивших в соответствии с «ключом». Понятие «трудность» чаще всего используется в отношении так называемых тестов достижений и интеллектуальных тестов, которые, как правило, с помощью мыслительных задач определяют уровень развития тех или иных интеллектуальных функций и в которых ответ в соответствии с «ключом» и является критерием правильности решения задания. В этих случаях трудность задания отражает процентное соотношение лиц, решивших или не решивших данное задание. В том случае, если задание выполняется 100 процентами испытуемых, трудность задания будет равна 0, если задание не выполняется ни одним испытуемым, то трудность данного задания будет равна 100. Математически такое процентное соотношение выражается в виде индекса трудности, который может колебаться от 0 до 1. В том случае, если трудность задания равна 0, то и индекс трудности будет равен 0, а если трудность равна 100, то индекс трудности будет равен 1. При выполнении задания 50 процентами испы-

туемых (трудность равна 50), индекс трудности соответственно принимает значение 0,5. Предполагается, что в том случае, если индекс трудности заданий слишком высок, то это означает, что наибольшая часть испытуемых не может правильно решить данное задание, поэтому оно не является диагностичным. Однако, с другой стороны, если индекс трудности слишком низок, это означает, что подавляющее большинство испытуемых правильно решают данное задание, что, в свою очередь, также говорит о его диагностической непригодности.

В тестах, диагностирующих не интеллектуальные, а личностные особенности индекс трудности также отражает процентное соотношение лиц, ответивших в соответствии с «ключом». При этом, естественно, не предполагается, что в таких тестах могут существовать правильные или неправильные ответы. В личностных тестах ответы в соответствии с «ключом» определяют выраженность у испытуемого того параметра личности, который является диагностическим конструктом данного теста. Например, если на задание, сформулированное как: *«Любите ли Вы бывать в шумных компаниях?»*, которое диагностирует экстравертированность личности, 90% испытуемых отвели «Да», то фактически это означает, что 90% окружающих нас людей являются экстравертами, что на самом деле неверно, а значит, проблема заключается в самом задании. Скорее всего, оно сформулировано таким образом, что провоцирует большинство испытуемых отвечать именно так, а не иначе.

Наиболее удачными с точки зрения трудности являются задания, индекс трудности которых составляет 0,5. Однако на основании закона нормального распределения, существует диапазон приемлемых значений индекса трудности, который находится в пределах от 0,16 до 0,84, или в процентном соотношении в соответствии с «ключом» на данный пункт должны отвечать не менее 16-ти процентов и не более 84-х процентов испытуемых. Указанный диапазон приемлемых значений не является случайным. Если предположить, что частоты распределения оценок испытуемых подчиняются закону нормального распределения, то в пределах $M \pm 1\sigma$ находятся около 68% всех испытуемых, а за этими пределами – по 16% испытуемых. Эти величины и прини-

маются за пороговые при отборе заданий из первичной формы теста.

Для того чтобы вычислить индекс для дихотомической шкалы ответов, достаточно лишь соотнести процентную долю испытуемых, давших на данное задание ответ, соответствующий «ключу», с процентной долей испытуемых, давших не соответствующий «ключу» ответ. Чем выше будет доля лиц, ответивших в соответствии с «ключом», тем ниже будет индекс трудности. Например, индекс трудности задания, на которое 80% испытуемых дали ответ, соответствующий ключу, будет равен 0,20.

Для вычисления индекса трудности для многоступенчатых шкал ответов обычно используется следующая формула:

$$Итр = \frac{Nb}{a \times n},$$

где: a – максимальный балл по шкале ответов, Nb – сумма баллов всех испытуемых по данному заданию, n – общее количество испытуемых

3. Характер распределения тестовых оценок. Данный статистический показатель используется при проверке тестовых заданий, которые предполагают многоступенчатую порядковую шкалу ответов. Цель определения данного показателя состоит в том, чтобы в окончательную форму теста попали бы только те задания, тестовые оценки по которым подчиняются закону нормального распределения, т.е. наибольшее количество испытуемых выбирает средние варианты ответа, в то время как самые крайние варианты ответов выбираются наименьшим количеством испытуемых. В том случае, если характер распределения тестовых оценок не соответствует закону нормального распределения, можно говорить о том, что данное задание теста не является достаточно диагностичным и «провоцирует» большинство испытуемых давать варианты ответов, относящиеся к крайним полюсам шкалы ответов.

Вопросы для самопроверки

1. В каких пределах могут колебаться значения коэффициента дискриминативности?
2. В каких пределах могут колебаться значения индекса трудности?
3. Почему приемлемые значения индекса трудности находятся в диапазоне 0,16 – 0,84?

5.5. Надежность теста

Одним из важных аспектов проблемы измерения в психологии является вопрос о степени доверия к его точности и стабильности. Иными словами, мы сталкиваемся с проблемой надежности измерения. Предположим, человек бросает мяч в баскетбольную корзину. В идеальных условиях его результативность будет зависеть только от его способности точно попадать в корзину. В реальных же условиях на его результативность может влиять большое количество факторов, например погодные условия, поведение болельщиков и т.п. В этих условиях мяч может либо не долететь до корзины, либо перелететь.

Таким образом, точность попадания в корзину будет зависеть от истинной способности баскетболиста поражать цель и от воздействия случайных факторов. При сравнении результатов точности попадания разных спортсменов можно предположить, что дисперсия результатов попадания может быть объяснена не только дисперсией «истинной», обусловленной различиями в выраженности психологического качества, но и разбросом, вызванным действием случайных факторов, т.е. дисперсией «погрешности»:

$$D_x = D_{ист.} + D_{погр.}$$

Отношение дисперсии истинной к дисперсии совокупной и будет называться **коэффициентом надежности измерения.**

Источником дисперсии «погрешности» являются разного рода ошибки. Появление ошибок может быть вызвано разными причинами, среди которых можно выделить две: неточность измерительного инструмента и условия опыта. С этих позиций выводят три вида надежности измерительного инструмента:

1. Надежность как устойчивость измерений. Этот вид надежности демонстрирует стабильность результатов во времени, т.е. измерение, произведенное одним и тем же измерительным инструментом в разное время, дает одинаковые результаты. При этом результаты измерения зависят как от самой измерительной процедуры, так и от условий измерения.

2. Надежность как точность измерений. Содержание этого вида надежности исчерпывается только качеством измерительного инструмента. Влияние условий эксперимента в этом случае невелико.

3. Надежность относительно постоянства условий. В этом случае надежность измерения может зависеть от влияния различных неконтролируемых факторов, которые необходимо учитывать.

Данные положения общей теории надежности измерения нашли свое конкретное воплощение и при оценке надежности теста как измерительного инструмента.

Под надежностью теста, в самом широком смысле, понимается характеристика того, в какой степени выявленные у испытуемых различия по тестовым результатам являются отражением действительных различий в измеряемых свойствах, а в каком они могут быть приписаны действию случайных факторов [4].

Распределение оценок испытуемых при выполнении теста, измеряющего одно качество, в идеальном случае совпадает с нормальным, и дисперсия при этом будет «истинной», т.е. будет отражать вариативность только измеряемого признака. Каждый испытуемый будет занимать определенное место по оценкам теста и теоретически это место для каждого члена выборки будет постоянным. При повторном проведении теста на той же самой выборке теоретически распределение мест по шкале оценок должно быть идентично первому. В данном случае мы можем говорить о тесте как о точном и надежном инструменте.

В реальных условиях повторное проведение теста дает иное распределение мест. При этом дисперсия нового распределения будет отличаться от дисперсии предыдущего на величину дисперсии ошибки. Сказанное можно выразить формулой, описываю-

щей надежность теста как отношение «истинной» дисперсии к «реальной», эмпирической.

$$R_t = \frac{D_t}{D_x},$$

где R_t – коэффициент надежности теста, D_t – «истинная» дисперсия, D_x – реальная дисперсия. Величина R_t служит основным показателем точности и устойчивости тестовых измерений и называется коэффициентом надежности теста.

Рассмотренные выше основные виды надежности измерительного инструмента нашли свое конкретное воплощение и в психометрике.

5.5.1. Надежность как устойчивость результатов

Надежность как устойчивость результатов во времени представляет собой согласованность результатов теста, получаемых при первичном и вторичном его применении по отношению к одним и тем же испытуемым в разные моменты времени. Коэффициентом надежности в этом случае будет коэффициент корреляции между результатами первого и повторного тестирования, определяемый по формуле К. Пирсона [17].

Данный вид надежности показывает, в какой степени результаты тестирования можно распространить на различные случаи его применения. В связи с использованием процедуры повторного тестирования одной и той же выборки данный вид надежности получил в психодиагностике название **ретестовой надежности**.

Повышение надежности в этом случае свидетельствует о повышенной чувствительности результатов тестирования к изменению состояния испытуемых и обстановки тестирования.

При определении ретестовой надежности особое значение имеет интервал между первичным и повторным тестированием. Слишком большой временной интервал может привести к снижению коэффициента надежности, поскольку повышается вероятность воздействия посторонних факторов (возрастные изменения измеряемых свойств, различные события, могущие повлиять на состояние испытуемых и их личностные особенности). Слишком маленький интервал времени будет способствовать припомина-

нию результатов первичного тестирования, что, в свою очередь, приведет к искусственному повышению коэффициента надежности. В психометрике традиционная точка зрения определяет оптимальный интервал для ретестовой надежности – 6 месяцев для взрослых и 3 месяца для детей [2], [4], [7]. Использование более длительных интервалов возможно, особенно при проверке некоторых видов валидности. Однако в этом случае необходимо фиксирование всех событий, произошедших за этот период с испытуемыми, и степень их влияния на тестируемую характеристику.

Вычисление ретестовой надежности представляет собой довольно простую процедуру, однако, наряду с этим достоинством, данный вид надежности имеет и существенные недостатки.

Первый из них связан с влиянием тренировки на результаты теста. При слишком коротком интервале времени между тестированиями у испытуемых может сформироваться навык работы с данной методикой, что приводит к улучшению результатов у отдельных испытуемых и соответственно к перестановке мест в ранговой последовательности, а как следствие – к снижению коэффициента надежности. Данный недостаток наиболее отчетливо выступает при проверке интеллектуальных тестов, тестов способностей и достижений. Для устранения влияния этого фактора используется предварительная тренировка всех испытуемых для формирования устойчивого навыка.

Однако такой способ, в свою очередь, увеличивает количество повторных проб и повышает вероятность появления второго, более существенного недостатка ретестовой надежности – припоминания результатов предыдущего тестирования.

Припоминание результатов обычно встречается при использовании небольших временных интервалов. В этом случае результаты двух обследований не будут независимы друг от друга, и коэффициент корреляции между результатами первичного и повторного тестирования будет искусственно завышенным. Увеличение интервала может привести к появлению неслучайных и долговременных изменений в измеряемых свойствах, что будет противоречить основной идее проверки надежности как характеристики теста.

Кроме того, установлено, что для большинства тестов способностей характерно улучшение показателей ретестовой надежности с возрастом испытуемых, что связано с замедлением темпа развития способностей и улучшением контроля условий выполнения теста. В этой связи необходимо измерять ретестовую надежность для разных возрастных групп, что особенно существенно для методик, ориентированных на широкий возрастной диапазон (Тест структуры интеллекта Д. Векслера, Шкала умственного развития Бине – Симона и т.п.).

Указанные выше особенности ретестовой надежности делают ее пригодной в качестве эффективного способа проверки для ограниченного числа тестов. К их числу относятся либо методики, где фактор припоминания не имеет значения (психофизические и сенсомоторные пробы, тесты на скорость), либо тесты с большим количеством пунктов (многофакторные личностные опросники).

5.5.2. Надежность параллельных форм

Другим видом надежности теста является надежность параллельных форм. Этот способ определения надежности психодиагностических методик предполагает использование двух взаимозаменяемых форм теста. Некоторые авторы относят этот способ определения надежности к разновидности ретестовой надежности [12]. Наиболее традиционной точкой зрения, которой придерживаемся и мы, предполагает рассмотрение этого способа определения надежности как самостоятельного и промежуточного между надежностью как временной устойчивостью и надежностью как точностью результатов теста [2], [4], [7].

Сущность этого способа заключается в обследовании выборки сначала основным набором заданий, а затем – дополнительным, аналогичным первому.

Другим путем проверки этого вида надежности является разбиение всей выборки на две примерно равные группы. Первичное тестирование заключается в тестировании первой группы одной формой теста (А), а второй группы – аналогичной формой (В). После небольшого интервала времени (примерно около недели) проводится повторное тестирование в обратном порядке.

Преимущество данного способа проверки надежности заключается, с одной стороны, в уменьшении влияния факторов тренировки и припоминания, характерных для ретестовой надежности и существенно снижающих эффективность ее применения. С другой стороны, вычисление надежности теста указанным способом приводит к возможности уменьшения временного интервала, что делает процедуру проверки более экономичной.

Вместе с тем, использование надежности параллельных форм имеет и свои недостатки. Основная трудность, как справедливо указывает А. Анастази, заключается в том, чтобы при конструировании теста «они (формы теста) на самом деле были параллельными» [2, с. 113]. Это означает, что обе формы теста должны конструироваться как два независимых теста, отвечающих одним и тем же требованиям (одинаковое число однотипных по содержанию и форме заданий, форма подачи и содержание инструкции, временные рамки, формы бланков, уровень трудности и дискриминативности). Естественно, что такая работа требует гораздо больших временных, интеллектуальных и энергетических затрат и существенно ограничивает сферу применения этого способа проверки надежности[7].

Несмотря на преимущества по сравнению с ретестовой надежностью, использование надежности параллельных форм не устраняет полностью влияния факторов тренировки и припоминания результатов, а лишь уменьшает их действие.

Наличие указанных недостатков ограничивает применение данного способа оценки надежности теста и заставляет исследователей обращаться к другим методам ее определения.

5.5.3. Надежность как точность измерений

Данный способ оценки надежности теста предполагает наличие внутренней согласованности между заданиями теста и их «суммарную работу на общую цель – точное измерение тестируемой характеристики». Процедура проверки предполагает вы-

[7] В частности, адаптация личностной методики KUD, созданной чешскими психологами, проведенная на кафедре психологии труда ЯрГУ, привела к созданию двух независимо друг от друга сделанных форм [18].

числение интеркорреляционных связей между заданиями, составляющими тест.

Наиболее простым и распространенным видом определения данного типа надежности является *метод расщепления*. Суть данного метода заключается в тестировании испытуемого двумя половинками теста.

Существуют несколько способов расщепления теста пополам:

Первый способ самый простой и заключается в разделении теста на четные и нечетные задания. Применение данной процедуры возможно только при равенстве показателей трудности в обеих половинках.

При наличии пунктов теста разных по степени трудности и дискриминативности применяется более трудоемкая процедура подбора половинок теста по приблизительной эквивалентности этих показателей.

Коэффициентом надежности теста в данном случае будет являться коэффициент корреляции между частями теста. Вместе с тем, следует оговориться, что данный коэффициент надежности будет отражать надежность лишь половины теста. Для вычисления коэффициента надежности всего теста используется формула Спирмена –Брауна:

$$R_t = \frac{2r_t}{1 + r_t},$$

где R_t – коэффициент надежности всего теста; r_t – коэффициент надежности половинки теста.

Данная формула может быть использована только в случае равных стандартных отклонений обеих половинок теста (σx_1 = σx_2). В случае их неравенства используется формула Фланагана:

$$R_t = \frac{4\sigma_1 \times \sigma_2 \times r_t}{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \times r_t}.$$

Для малых выборок предлагается расчет этого же показателя по формуле Кристофа [4], [7].

$$R_t = \frac{2}{n-1} + \frac{n-3}{n-1} \cdot \frac{4\sigma_1 \times \sigma_2 \times r_t}{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \times r_t},$$

где R_t – коэффициент надежности всего теста; r_t – коэффициент надежности половинки теста; n – число заданий; σ_1 и σ_2 – стандартные отклонения одной и второй половинки теста соответственно.

Для вычисления коэффициента надежности всего теста может использоваться формула Рюлона, основанная на соотношении дисперсий:

$$R_t = 1 - \frac{D_\Delta}{D_x},$$

где D_Δ – дисперсия разностей между результатами каждого испытуемого по двум половинкам теста; D_x – дисперсия суммарных результатов.

При расщеплении тестов на скорость используется специальная процедура. Сначала определяется минимальное время выполнения теста, затем отсчитывается половина и четверть от этого временного показателя. После этого испытуемым предлагается выполнять тест. По истечении половины отсчитанного прежде временного интервала они, по команде экспериментатора, ставят отметку на бланке и продолжают работать дальше до истечения еще четверти эталонного интервала. Коэффициентом надежности в данном случае является коэффициент корреляции между числом заданий, решенных до первого сигнала экспериментатора (половина минимального интервала), и числом задач, решенных между первым и вторым сигналом экспериментатора (четверть эталонного интервала) [4], [7].

Расщепление теста на две половинки является наиболее простым способом определения надежности как внутренней согласованности. Если теоретически предположить, что мы можем расщеплять тест на сколько угодно частей, то конечным результатом будет расщепление теста до такой степени, когда каждая его часть будет представлять одно задание. Эквивалентность частей теста в этом случае подразумевает, что задания теста должны быть однородны по содержанию и трудности. Данный способ определения надежности как внутренней согласованности часто называют **надежность по однородности.**

Наиболее простым способом определения надежности по однородности является метод Кьюдера –Ричардсона, имеющий несколько вариантов. Наиболее простым вариантом является следующий:

$$R_t = \frac{K}{K-1} \times \left(\frac{\sigma_X^2 - \sum pq}{\sigma_X^2} \right),$$

где K – число заданий в тесте; p – индекс трудности, выраженный в виде доли $\frac{ИТ}{100}$, q = 1–p, σ_x^2 – дисперсия первичных результатов теста.

Предложенная формула имеет ограниченную сферу применения и используется при дихотомической шкале, когда оценка результатов производится по принципу «все или ничего» («решил / не решил»; «верно / не верно»; «да / нет» и т.п.). Кроме того, данная разновидность формулы Кьюдера – Ричардсона применяется в случае неучета коэффициентов дискриминативности.

При многоступенчатой шкале используется коэффициент «Альфа», предложенный Л. Кронбахом:

$$R_t = \frac{n\left(\sigma_X^2 - \sum \sigma_{Xi}^2\right)}{(n-1)\,\sigma_X^2}$$

где $\sum \sigma_{Xi}^2$ – сумма дисперсий балльных оценок по каждому заданию; σ_X^2 – дисперсия суммарных показателей теста; n – число заданий в тесте.

5.5.4. Надежность относительно постоянства условий

Надежность относительно постоянства условий в психодиагностике проявляется в **надежности экспериментатора.** В любом исследовании экспериментатор может выполнять три функции: организатора исследования, оценщика и интерпретатора.

Постоянство экспериментатора как организатора исследования предполагает, что его личностные особенности, социальное положение и другие характеристики не оказывают существенного влияния на обследуемых лиц и соответственно на результаты тес-

тирования. Достигается это за счет стандартизации тестового материала, инструкций, условий проведения тестирования.

Постоянство подхода оценщика касается ведения протоколов и подсчета результатов теста. При наличии стандартных бланков ответов, ключей объективность оценки повышается, что приводит и к повышению надежности.

Постоянство подхода экспериментатора как интерпретатора обеспечивается, если одинаковые оценочные данные, полученные от испытуемых, приводят к одинаковым заключениям. В психометрике такое положение обеспечивается наличием норм, стандартных профилей оценки или критериев.

Большинство тестов, используемых для массовых обследований, достаточно жестко стандартизированы, поэтому их проведение, обработка и интерпретация сводят дисперсию ошибок, вызванных влиянием экспериментатора, практически к нулю. Однако некоторые типы тестов, особенно используемые в клинической практике (тесты креативности и проективные тесты), оставляют много свободы исследователю, особенно в части интерпретации полученных результатов. Это существенно повышает роль надежности относительно экспериментатора по сравнению с более традиционными ее видами.

В практике психометрической проверки тестов обязательным условием является проверка надежности как устойчивости и как точности результатов. Наиболее популярными способами является проверка ретестовой надежности и надежности по однородности.

Коэффициент надежности. Поскольку во всех видах определения надежности в качестве ее коэффициента выступает коэффициент корреляции, то уровень его значимости определяется принятым в математической статистике путем и будет зависеть от объема выборки. Коэффициент надежности как коэффициент корреляции может изменяться в пределах от –1 до +1, однако тест считается надежным, когда его коэффициент находится в пределах 0,6 – 1. Случаи, где коэффициент надежности меньше, чем 0,6, должны рассматриваться особо [7]. Некоторые авторы, вместе с тем, считают, что для большинства известных методик редко удается получить коэффициент надежности выше 0,7 – 0,8.

Коэффициент надежности, как и любой статистический показатель, имеет доверительный интервал, в котором он может оказаться при его вычислении для генеральной совокупности. Для коэффициента надежности он равен:

$$CL_z = Z\,(r) \pm Z\text{крит.}\ \sigma_{rtt},$$

где r – коэффициент надежности, σ_{rtt} – стандартная ошибка коэффициента надежности. Стандартная ошибка высчитывается по формуле:

$$\sigma_{rtt} = \frac{1}{\sqrt{n-3}}.$$

На практике значение имеет только нижняя граница доверительного интервала коэффициента надежности, которая равна для выборок больше 30 человек: $R_t = 1{,}96\sigma_{rtt}$. Превышение коэффициента надежности нижней границы доверительного интервала означает, что тест надежен и при неблагоприятных условиях.

Величина коэффициента надежности подвержена влиянию самых разнообразных факторов, среди которых ведущее значение имеет количественный и качественный состав выборки. Традиционно считается, что увеличение объема выборки ведет к снижению погрешности. Поэтому выборки для вычисления надежности должны быть достаточно большими. Минимальный объем выборки для расчета надежности должен быть не меньше 150 – 200 человек. Качественный состав выборки должен полностью отражать генеральную совокупность.

Важным свойством выборки, влияющим на надежность теста, является диапазон индивидуальных различий в группе. Если уровень выраженности какой-либо способности (скорость чтения) у всех членов группы будет одинаковым, то, вероятнее всего, корреляция этой способности с любыми другими будет очень низкой. Отсюда предсказать положение индивида по какой-либо другой способности на основе знания его показателей о скорости чтения будет практически невозможно [2].

Изменчивость выборки и степень ее однородности является важным фактором, воздействующим на коэффициент надежности. В частности, если при создании теста для определения его надежности была использована выборка школьников 5 – 11 клас-

сов, то нельзя полагать, что использование его на выборке девятиклассников даст высокие результаты. В случае использования теста на более однородной выборке, чем выборка стандартизации, коэффициент надежности необходимо рассчитывать заново.

Коэффициент надежности может изменяться не только от степени индивидуальных различий в выборке, но и от среднего уровня измеряемой способности [2]. Поэтому при конструировании тестов часто используется прием разбиения выборки стандартизации на более однородные подгруппы по разным признакам (полу, возрасту, уровню образования и т.п.).

5.5.5. Основной психометрический парадокс

В психометрике существует феномен, отражающий взаимоотношения между дискриминативностью заданий теста и ретестовой надежностью, который получил название основного психометрического парадокса. Сущность этого феномена заключается в том, что задания, обладающие хорошей различительной способностью, при повторном тестировании характеризуются нестабильностью ответов, что приводит к снижению показателя ретестовой надежности теста. И наоборот, ответы на задания, имеющие низкую степень дискриминативности, при повторном тестировании оказываются более стабильными. Психометрический парадокс проявляется при конструировании личностных опросников [3, 4]. Например, в работах зарубежных психологов в 1940-х годах было показано, что вопросы, хорошо дифференцирующие больных неврозом от других категорий больных или здоровых испытуемых, при повторном тестировании оказывались ненадежными, т.е. вероятность получения того же самого ответа была очень мала. С другой стороны, вопросы, которые давали стабильно одинаковый ответ в обоих тестированиях, плохо дифференцировали испытуемых на группы с высокой и низкой степенью выраженности диагностируемого психического качества [4].

Механизм возникновения парадокса, по мнению одного из известных его исследователей М. Новаковской, кроется в содержании инструментария, используемого в гуманитарных науках, а именно в неоднозначности понимания формально неизменных утверждений и различиях в выраженности психологических черт

[3]. С точки зрения автора, причинами, лежащими в основе того или иного ответа, могут быть: а) выраженность черты у обследуемого; б) значение, придаваемое вопросу; в) степень легкости в принятии решения об ответе. Исходя из этого, автором предлагается различать два типа психометрического парадокса: парадокс типа А и парадокс типа В [3].

Первый тип парадокса возникает, когда вопросы поддаются различному истолкованию или когда трудно принять решение об ответе. В этом случае утверждения обладают хорошей дискриминативной силой, но и значительной вариабельностью понимания. Например, «Ваше настроение обычно плохое».

Парадокс типа В в основном характерен для утверждений, однозначно понимаемых подавляющим большинством испытуемых, где легко подобрать ответ. В этом случае утверждение будет обладать низкой степенью дискриминативности, но при этом будет характеризоваться стабильностью ответа при повторном тестировании (высокой надежностью). Например, вопрос, на который можно легко дать однозначный ответ: «Вы курите?».

По мнению М. Новаковской, практически все вопросы могут вызывать парадокс любого типа, однако возможны и «идеальные» случаи, которые достигаются путем конкретизации проявления диагностируемой черты. Примером такого задания может быть утверждение: «Я часто чувствую себя по утрам бодрым и полным сил», которое характеризуется высокой степенью однозначности понимания содержания утверждения, с одной стороны, и высокой дискриминативностью, с другой стороны. Знание о существовании психометрического парадокса позволит создателю теста более внимательно подойти к формулированию содержания утверждений.

Надежность является важной характеристикой теста, но сама по себе ценности не имеет. Поэтому следует помнить, что ее ценность состоит в том, что она необходима для достижения другой важной характеристики теста – его валидности.

Вопросы для самопроверки

1. В чем смысл понятия «погрешность измерения»?
2. Каковы основные виды надежности теста?
3. В каких значениях могут колебаться значения коэффициента ретестовой надежности?
4. В каких случаях чаще всего понижается надежность теста?
5. В чем смысл метода расщепления?
6. Что такое «психометрический парадокс»?

5.6. Валидность теста

В психологической диагностике валидность – это обязательная и наиболее важная часть информации о психологическом тесте. В самом общем смысле валидность указывает нам на то, что тест измеряет и насколько хорошо он это делает. Исходя из этого, вопросы валидности – это вопросы о том, какие и насколько обоснованные выводы могут быть сделаны на основе тестовых оценок. Все вопросы, относящиеся к валидности, могут быть сведены к двум основным: 1) какие выводы можно сделать о том, что измерялось тестом?; 2) какие выводы могут быть сделаны о реальном поведении человека на основе тестовых оценок?

1-й вопрос выясняет внутреннюю природу самого измерения, поскольку тест как измерительный инструмент является операционализацией какого-либо психического свойства. В этом плане существенным вопросом является вопрос о том, в какой степени тестовые баллы отражают содержание данного психического свойства, и в этом смысле следует говорить о внутренней валидности.

2-м вопросом выясняется полезность тестового измерения как предсказателя поведения. В этом плане существенным вопросом является вопрос о том, насколько тестовые баллы связаны с жизненными проявлениями изучаемого свойства, и в этом смысле следует говорить о внешней валидности.

Необходимо отметить, что валидность не измеряется, о ней можно только судить по набору различных коэффициентов ва-

лидности. То есть валидность есть нечто такое, о чем судят как об удовлетворительном или неудовлетворительном для вынесения соответствующего суждения.

Помимо этого, понятие валидности относится не столько к самому тесту, сколько к цели, ситуации и процедуре его применения. Нельзя задать вопрос «Это валидный тест?», следует спросить «Насколько валиден этот тест для принятия какого-либо решения?» или «Для каких целей этот тест является валидным?». Например, тест, являющийся достаточно валидным для определения каких-либо психических отклонений, может быть абсолютно невалидным для определения успешности выполнения какой-либо профессиональной деятельности. Следовательно, тест не может иметь высокую или низкую валидность вообще, он может быть валидным или невалидным лишь по отношению к определенной цели применения.

В понятие валидности входит большое количество самой разнообразной информации о тесте. Различные категории этих сведений и способы их получения образуют виды валидности: критериальную, конструктную, содержательную, синтетическую, экологическую, очевидную и др.

5.6.1. Критериальная валидность

Критериальная валидность – это наиболее часто используемый вид валидности. Критериальная валидность отражает соответствие диагноза или прогноза, определяемого на основе тестовых оценок определенному кругу жизненных критериев измеряемого качества. Критериальная валидность измеряется путем корреляции баллов по тесту и показателей по внешнему параметру, избранному в качестве критерия, при этом это должен быть именно внешний, т.е. жизненный, независимый от тестирования критерий. Критериальная валидность делится на ***текущую и прогностическую.***

Текущая валидность. В разных литературных источниках ее называют также совпадающей или конкурентной [12], [61]. Текущая валидность – это характеристика теста, отражающая его способность различать испытуемых на основании диагностического признака, являющегося объектом исследования в данной методике

(уровень способностей, вербальный интеллект, тревожность и т.д.). В более узком значении текущая валидность – это установление соответствия результатов валидизируемого теста независимому критерию, отражающему состояние исследуемого тестом качества в момент проведения исследования. В качестве критерия может выступать какой-либо показатель деятельности, экспертная оценка и т.д. Например, в качестве внешнего критерия для валидизации теста, предназначенного для оценки уровня интеллектуального развития школьников, может выступать уровень школьной успеваемости. При этом мы исходим из предположения о том, что уровень школьной успеваемости школьника напрямую связан с уровнем его интеллектуального развития. Процедура определения валидности в данном случае проста. Она заключается в подсчете коэффициета корреляции между тестовыми баллами, полученными по валидизируемому тесту, и оценками по школьным предметам в одной и той же группе испытуемых.

Текущая валидность является одной из ведущих характеристик валидности любой психодиагностической методики.

Прогностическая валидность характеризует степень точности и обоснованности суждений о диагностируемом качестве по его результату, спустя определенное время после его измерения. Исследователь проводит тест, делает определенные предсказания на основе тестовых оценок, предпринимает определенные действия, исходя из этих предсказаний, и ждет, что произойдет. Например, валидизируя тест на определение способностей к педагогической деятельности, исследователь должен, прежде всего, протестировать некую группу испытуемых, планирующих заниматься данным видом деятельности, и на основе тестовых оценок сделать прогноз относительно успешности выполнения данного вида деятельности для отдельных испытуемых. После того, как данная группа испытуемых проработает определенное время по педагогической профессии, исследователь должен получить объективную, жизненную информацию относительно успешности выполнения каждым испытуемым деятельности педагога и определить, насколько сбылся его прогноз относительно отдельных испытуемых. Математическим показателем валидности в данном случае будет являться коэффициент корреляции между тестовы-

ми оценками и объективными показателями успешности выполнения деятельности. В том случае, если он будет достаточно значимым, мы сможем говорить о том, что на основе тестовых оценок можно действительно предсказать успешность выполнения деятельности, а значит, наш тест обладает достаточной прогностической валидностью для решения такого рода задач. Таким образом, прогностическая валидность непосредственно указывает на предсказательную силу теста и определяет степень обоснованности прогноза.

Отметим, что текущая и прогностическая валидность подразумевают разные цели применения методики: текущий диагноз, оценка настоящего положения, с одной стороны, и прогнозирование развития качества или успеха в деятельности – с другой. Оценка прогностической валидности является более сложной проблемой вследствие временной отдаленности тестирования от реализации прогноза. Особенно сложно определить прогностическую валидность для тестов, направленных на измерение сложных психологических конструктов, таких как общие способности, особенности личности, уровни достижений и т.д. Развитие этих психологических свойств в большой степени зависит от приобретаемых знаний и навыков и может резко изменяться в зависимости от обстоятельств жизни и деятельности индивида. Иногда требуется глубокий анализ психологических конструктов, лежащих в основе тестов, чтобы понять динамику их развития. Это сближает прогностическую критериальную валидность с конструктной валидностью.

В эмпирических исследованиях критериальной валидности наибольшая трудность заключается в поиске соответствующей ***критериальной меры.*** Несмотря на то, что в качестве критериальной меры может выступать любая оценка поведения, практически подобрать соответствующую критериальную меру бывает достаточно сложно.

Все разнообразие критериев валидизации обычно сводится к нескольким типам. Приведем здесь две такие классификации.

Дж. Тиффин и Э. МакКормик рассматривают критерии, применяемые в индустриальной психологии, и объединяют их в 4 класса: критерии исполнительской природы, субъективной при-

роды, физиологической природы и (в особых обстоятельствах) несчастные случаи [69].

1. Критерии исполнительской природы имеют ярко выраженные поведенческие характеристики. В индустриальной психологии они обычно представляются в виде результативности выполнения работы, как, например, количество единиц изготовленной продукции. В некоторых случаях используются другие аспекты исполнения, например, мера визуального различения (при разных уровнях освещения) или время реакции (на различные дорожные сигналы).

2. Субъективные критерии охватывают различные виды ответов, которые отражают субъективные реакции людей (установки, мнения, суждения, предпочтения и т.п.). Обычно их получают методами интервью, опросников и рейтинговых оценок, даваемых экспертами. Часто (например, в случае опросников установок) ответы даются на ряд вопросов или утверждений и затем переводятся в общую оценку. Наиболее распространённый тип субъективного критерия – рейтинги, например, рейтинговая оценка начальником своих подчинённых. Рейтинги, используемые в качестве критериев, могут составляться на основе «общего» выполнения работы или на основе какого-то частного аспекта исполнения. Хотя рейтинги субъективны по природе, иногда они могут быть наиболее подходящим методом получения некоторых показателей работы.

Основное требование, предъявляемое к экспертам, заключается в том, что они должны не только хорошо знать оцениваемое лицо, но и наблюдать его в различных ситуациях, в которых проявляется оцениваемая характеристика.

Количество экспертов может быть различным, однако наиболее надежной представляется оценка, даваемая группой экспертов (5 – 7 человек). Мнение одного эксперта, как правило, может использоваться в тех случаях, когда эксперт является высококомпетентным лицом и ведущим специалистом в своей области. Существует несколько способов подсчета экспертной оценки, среди которых наиболее надежным считается *метод средневзвешенной оценки.* Данный метод предполагает оценку каждого испытуемого по заранее установленной шкале несколькими экспертами не-

зависимо друг от друга, т.е. оценка, даваемая одним экспертом, неизвестна другим. Общий балл подсчитывается путем усреднения всех индивидуальных оценок экспертов. Данный метод считается более надежным и достоверным по сравнению с *методом коллективной оценки*, где группа экспертов совместно оценивает испытуемого по заранее установленной шкале, вырабатывая общее мнение о выраженности той или иной психологической характеристики. Недостатком указанного способа экспертной оценки является ее зависимость от индивидуально-личностных особенностей экспертов или их статуса в группе. В случае трудности соотнесения выраженности оцениваемого признака с его положением на шкале, наиболее часто проявляющейся при оценке малодифференцированных признаков, а также при невысокой квалификации экспертов может использоваться *метод ранжирования* или *метод попарного сравнения.*

Первый способ экспертной оценки предполагает использование вместо оценочной шкалы ранжирования, т.е. приписывания испытуемому определенного места по уровню выраженности качества. Например, ученик Иванов по степени честности занимает первое место в классе, ученик Петров – второе и т.д. Усредненной оценкой является не среднеарифметическая величина, а медиана.

Второй метод предполагает попарную расстановку объектов по альтернативным позициям. При этом каждый объект сравнивается с каждым по выбранному признаку. Например, эксперты сравнивают ученика Иванова по уровню выраженности честности попарно с каждым членом группы, отмечая, кто в паре честнее: Иванов или другой ученик. Показателем места, занимаемого объектом в ряду других, служит общее число его предпочтений по сравнению с остальными по данным всех экспертов. Достоинством данного метода, помимо указанных выше, также является отсутствие количественной оценки, что воспринимается экспертами как субъективно более легкая процедура.

Другие типы включают опросники мнений и отношений, например, получение мнений работников по поводу различных ситуативных переменных (поведение начальника, политика компании, условия работы).

3. Физиологические критерии включают такие меры, как сердечный ритм, кровяное давление, электрическое сопротивление кожи и т.п. Обычно такие критерии используются при изучении влияния окружающей среды или других ситуативных факторов.

4. Количество несчастных случаев можно считать критерием валидизации тогда, когда стоит цель изучения правил безопасности. Такой критерий может использоваться в двух обстоятельствах. Во-первых, при оценке относительной безопасности двух или более ситуаций, как, например, различные типы оборудования, различные методы работы, условия работы и т.д. Во-вторых, при отборе на работу персонала, который должен работать без несчастных случаев; малый опыт несчастных случаев может быть критерием отбора нужных людей.

Выяснив, какими критериями в данной области он располагает, исследователь приступает к их отбору. При этом надо учитывать, что хороший критерий должен удовлетворять определённым требованиям. Он должен быть релевантным, незагрязнённым и надёжным.

1. Релевантность критерия. Это степень, в которой критериальная мера значима для целей измерения. В области индустриальной психологии критерий является релевантным, если оценка работника по критерию определяется теми же факторами, которые определяют успех в работе. Каждая работа выполняется с определённой целью или рядом целей; релевантность, таким образом, является адекватностью критерия как показателя способностей индивида в выполнении этих целей. У большинства видов работ имеется несколько целей, и поэтому встаёт вопрос, что лучше – использовать один общий критерий или несколько отдельных критериев (субкритериев) для отдельных целей.

Теоретически существует основной критерий, который служит основанием для характеристики поведения индивида. Обычно основной критерий рассматривается в контексте общего выполнения, как, например, «общая ценность человека для компании – в конечном итоге». Предполагается, что понятие основного критерия может применяться к каждой из разных целей работы по мере того, как эта цель побуждает общее исполнение. Так как эталон основного критерия скорее теоретический, чем реальный,

чем более реальный критерий приближается к эталону, основному стандарту, тем более релевантным он является.

На практике этот принцип реализуется следующим образом: подбираются критерии, находящиеся в определённой зоне приближения к главной цели оценки. В зависимости от близости к основному критерию такие критериальные меры могут быть прямыми и промежуточными.

2. Свобода от «загрязнения» (контаминации). Выполнение индивидом работы является функцией как переменных индивида, так и переменных ситуации. До этого мы рассматривали критерии, которые отражают индивидуальные различия, в то время как любое влияние различий в ситуативных переменных может служить «загрязнению» индивидуальной критериальной меры. Поясним на примерах. В качестве критериальной меры для валидизации теста на определение способностей к какой-либо рабочей специальности мы можем избрать производительность труда. Представим двух рабочих, обслуживающих ткацкие станки, но при этом каждый из них обслуживает свой тип станка. В данной ситуации производительность труда этих рабочих может отражать не их способности к выполнению данного вида деятельности, а чисто технические особенности их станков, и продуктивность их работы во многом будет зависеть от возможностей станка. В данном примере тип станка и будет являться фактором загрязнения критерия.

Выделяют два основных подхода к учёту «загрязнения» критерия. Один из них – использование в исследовании только тех индивидов, для которых все побочные переменные уравнены или почти уравнены; однако это может существенно ограничить количество испытуемых. Второй подход – составить определённую гипотезу в отношении влияния «загрязнённых» переменных, при этом обычно используется статистика. Например, если изучаются 10 рабочих, занятых на двух типах станков, продуктивность каждого рабочего можно соотнести со средней продуктивностью всех рабочих, занятых на этом типе станка.

Если исследователя интересуют критерии для сравнения не индивидуальных различий, а разных ситуаций (например, работы при различных уровнях освещения), то контаминация критерия

может произойти, если другие (неконтролируемые) переменные влияют на выполнение работы. Например, группа рабочих при одном уровне освещения работает на машинах, находящихся в лучшем состоянии, чем у рабочих другой группы. Контаминация может также произойти, если вмешаются индивидуальные переменные, например, одна группа рабочих может оказаться более способной, чем другая.

Приведенные выше примеры показывают проявление контаминации объективных критериев. Однако наиболее часто этот феномен имеет отношение к *субъективным* критериям валидности теста, в качестве которых выступают экспертные оценки специалистов. Контаминация субъективных критериев характеризуется влиянием результатов тестирования на мнения экспертов. Например, учитель, выступающий в роли эксперта при оценке выраженности какой-либо личностной характеристики, накануне случайно ознакомился с результатами тестирования учеников и запомнил оценки по тесту некоторых из них. На следующий день, выступая в роли эксперта, он ставит данному ученику соответствующий балл, ориентируясь не на свое собственное мнение о выраженности данной черты, а на результат тестирования. Во избежание подобных случаев психологу рекомендуется ограничить доступ к результатам тестирования посторонних лиц, а специалистов, выступающих в роли будущих экспертов, предупреждать о том, что до окончания проверки теста психолог не имеет права делать какие-либо выводы о конкретных испытуемых.

Отметим, что контаминация критерия, безусловно, является нежелательным фактором. Поэтому его влияние на получаемые результаты следует по возможности минимизировать, если его нельзя избежать совсем.

3. Надёжность критерия. Если критерий используется в качестве показателя исполнения работы индивидом, это должен быть стабильный показатель. Он должен отражать относительно постоянный, или непрерывный, уровень исполнения. Если исполнение по критерию меняется слишком часто, то не будет возможности найти тест, который сможет его предсказывать. Мера, которая сама по себе нестабильна, не может быть чем-либо

предсказана. Стабильность исполнения по критерию называется надёжностью критерия.

4. Практичность критерия – простота и невысокая стоимость получения соответствующих критериальных показателей.

5. Дискриминативность критерия. Данное требование заключается в том, что выбранный критерий должен хорошо дифференцировать людей с высоким и низким его значением. Например, выбранный в качестве критерия показатель скорости какой-либо производственной операции может оказаться не дискриминативным, поскольку сама операция является настолько простой, что через определенный промежуток времени ее осваивают все работники в одинаковой степени.

Среди различных способов определения надежности критерия очень распространен метод сравнения критериальных оценок за два отрезка времени.

Показателем критериальной валидности всегда выступает статистический коэффициент. Наиболее часто используются различные виды корреляционного анализа связи между индивидуальными оценками по тесту и оценками по критерию. В зависимости от конкретной ситуации применяется тот или иной коэффициент. Если оба ряда переменных представляют собой количественные континуумы, используется коэффициент корреляции произведений моментов Пирсона [7].

Если изучается связь качественных переменных, измеряемых в порядковых шкалах и шкалах наименований, то используется корреляция качественных признаков. В случае порядковых шкал (например, когда сравнивается степень устойчивости внимания испытуемого: очень устойчивое, устойчивое, лабильное, рассеянное – с его академической успеваемостью) применяется ранговая корреляция, рассчитываемая по коэффициентам Спирмена и Кендалла [42].

Если сравниваемые переменные измеряются в шкалах наименований (например, результат проективной методики сравнивается с медицинским диагнозом), то применяются коэффициенты сопряженности. Чаще всего пользуются коэффициентом сопряженности Пирсона [17].

Если требуется вычислить корреляцию признаков, измеряемых в дихотомической шкале наименований (в обоих рядах), то применяется упрощенный коэффициент корреляции произведений моментов Пирсона, чаще называемый коэффициентом ассоциации [18].

Когда одна из двух переменных измерена в дихотомической шкале наименований, а другая в интервальной шкале отношений или порядковой шкале (например, при сравнении порогов абсолютной чувствительности с принадлежностью к двум критическим группам), применяется коэффициент бисериальной корреляции.

Необходимо отметить, что отдельно вычисленный коэффициент валидности всегда носит приблизительный характер, и эта приблизительность обусловлена многими факторами. Во-первых, все условия валидизации невозможно определить полностью, и поэтому всегда остается множество неучтенных факторов. Во-вторых, логика критериальной валидизации предполагает также валидность самого критерия. Проверка такой валидности представляет собой весьма трудную проблему, и поэтому зачастую тесты валидизируются относительно наиболее доступного, а не наиболее соответствующего критерия. В-третьих, условия валидизации по критерию предполагают, что выборка полностью репрезентативна относительно той популяции, для которой позже будут делаться заключения. В реальной ситуации это требование осуществить трудно, и поэтому репрезентативность выборки валидизации всегда требует дополнительной проверки.

Наибольшая сложность интерпретации критерия валидности заключается в следующем. Критериями, как правило, выступают внешние для психологии социально-практические показатели. Поэтому в ходе валидизации решаются сразу две задачи: собственно задача измерения валидности методики и оценка пригодности самого критерия. Если получен значимый коэффициент корреляции, то можно считать, что обе задачи решены положительно. Но если корреляции не обнаружено, то остается неопределенность: либо невалидна сама методика, либо выбран неадекватный критерий. Например, при валидизации теста на стрессоустойчивость в работе оператора в качестве критериальной меры

может быть выбрано количество аварийных ситуаций в его работе. Если в данном случае коэффициент валидности оказывается незначимым, то это может говорить либо о том, что тестовый балл не отражает стрессоустойчивость оператора, либо неверна гипотеза о наличии связи между стрессоустойчивостью и количеством аварийных ситуаций.

5.6.2. Конструктная валидность

Конструктная валидность теста показывает, насколько его результаты могут рассматриваться в качестве меры некоего теоретического конструкта или свойства. Это один из основных видов валидности, отражающий степень репрезентации исследуемого психического свойства в результатах теста. С помощью конструктной валидности исследователь определяет область теоретической структуры психологических явлений, которые измеряет тест. Понятие конструктной валидности указывает на высокую зависимость эмпирических связей теста от теоретических представлений автора теста об измеряемом свойстве.

Смысл валидизации по конструкту заключается в том, чтобы выделить основное понятие, конструкт, на измерение которого направлен тест, поставить ему в соответствие набор эмпирических индикаторов (собственно тестовых заданий) и обосновать это соответствие. Именно в этом обосновании заключается сложность конструктной валидизации. Приступая к ней, исследователь может начать с самого теста, который он хочет лучше понять, а может – с идеи, которую он хочет оценить. Интерпретация теста, основанная на конструктной валидности, создается постепенно и, вероятно, никогда не заканчивается. По мере роста знаний создается более полный список влияний, воздействующих на тестовые оценки, и можно подсчитать силу и характер каждого из них. Как отмечает Л. Кронбах, в настоящее время интерпретация даже самых устоявшихся психологических тестов далека от идеала [56].

Проверка конструктной валидности может осуществляться следующими способами:

1. Сопоставление исследуемого теста с другими методиками, конструктное содержание которых известно. Это наиболее про-

стой для исследователя способ. Между баллами по двум тестам подсчитывается коэффициент корреляции. Положительная корреляция указывает на то, что разрабатываемый тест измеряет примерно ту же сферу поведения, способность, психологическое качество, что и эталонная методика. Такая процедура напоминает валидизацию по критерию, т.к. можно сказать, что в данном случае эталонный тест, валидность которого определена, выступает в качестве независимого критерия. В данном случае, в отличие от критериальной валидности, не требуется слишком высокого коэффициента корреляции. Если окажется, что новый и эталонный тесты практически идентичны по содержанию, и разрабатываемая методика не обладает преимуществом краткости или легкости применения, то это означает лишь дублирование эталонного теста.

2. Во многих случаях бывает трудно найти какой-либо апробированный тест на измеряемое свойство, валидность которого была бы известна. В этом случае исследователь формулирует ряд предсказательных гипотез о том, как будет коррелировать его новый тест с другими тестами, измеряющими родственные характеристики. Эти гипотезы выдвигаются на основании теоретических представлений об измеряемом свойстве. При этом конструктная валидность характеризуется не только связями проверяемого теста с близкородственными показателями, но и с теми, с которыми исходя из гипотезы значимых связей наблюдаться не должно. Эти подходы определяются как конвергентная и дивергентная валидность.

3. Особым и наиболее информативным способом проверки конструктной валидности является факторный анализ. Он позволяет статистически проанализировать структуру валидизируемого теста, определить наличие и влияние на результаты тестирования скрытых, неочевидных переменных. А. Анастази подчеркивает, что главное назначение факторного анализа состоит в упрощении описания данных путем сокращения числа категорий до нескольких общих факторов или свойств [2]. Исключительная важность процедуры факторного анализа является основанием для выделения ее в особый вид конструктной валидности – *факторную валидность*.

Необходимо отметить, что факторный анализ, как правило, применяется для оценки конструктной валидности многошкальных психодиагностических методик, т.е. тех методик, которые предполагают измерение не одной, а нескольких психологических характеристик. Проверка конструктной валидности с помощью факторного анализа подчиняется следующей логике. Первоначально исследователь на теоретическом уровне выстраивает общую структуру методики, т.е. определяет, какой пункт теста включается в ту или иную шкалу методики. Затем он проводит процедуру факторного анализа. В процессе факторного анализа прежде всего подсчитываются интеркорреляционные связи между всеми пунктами методики, т.е. подсчитываются корреляции каждого пункта с каждым другим пунктом, и те пункты теста, которые связаны друг с другом наиболее тесно, объединяются на основе этого в один фактор. Практическим результатом факторного анализа является «факторный вес» каждого пункта теста в отдельном факторе, т.е. мера включенности данного пункта в данный фактор. Таким образом, в результате этой процедуры исследователь снова получает информацию о том, в какую шкалу методики включается тот или иной пункт, но это будет уже информация, полученная на другом, эмпирическом уровне. Таким образом, в результате факторного анализа получается новая, эмпирическая структура методики, далеко не всегда полностью совпадающая со структурой теоретической, т.е. далеко не всегда тот или иной пункт теста, первоначально отнесенный исследователем к той или иной шкале методики, будет относиться к этой же шкале в результате факторного анализа. Отметим, что факторная валидность не имеет, как такового, единого коэффициента валидности. Показателями факторной валидности являются факторные веса отдельных пунктов в каждой шкале методики. Факторная валидность оценивается по степени совпадения или несовпадения теоретической и эмпирической структуры методики.

4. Внутренняя согласованность пунктов теста также является важным аспектом конструктной валидности. Она отражает то, насколько отдельные пункты подчинены основному направлению теста как целого и ориентированы на изучение одного и того же конструкта. В качестве критерия используется суммарный балл

данного теста, который коррелируется с ответами на каждое задание. Следует отметить, что критерий внутренней согласованности указывает лишь на меру связи всего содержания теста с измеряемым конструктом, не давая информации о природе измеряемого свойства, и поэтому одного лишь измерения внутренней согласованности недостаточно для утверждения о конструктной валидности теста.

5. Одним из способов проверки конструктной валидности теста является введение экспериментальной переменной. Данный способ заключается в первоначальном проведении тестирования в обычных для испытуемых условиях, а затем в ситуации действия фактора, который может существенно повлиять на результаты теста. Например, при валидизации теста диагностики тревожности мы проводим тестирование студентов в обычной ситуации и затем перед экзаменом. Коэффициентом валидности в этом случае будет степень достоверности различий между результатами первого и второго тестирования.

5.6.3. Содержательная валидность

Содержательная валидность – это третий важный вид валидности, в корне отличающийся и от критериальной, и от конструктной валидности.

Валидность по содержанию означает систематическую проверку содержания теста с тем, чтобы установить, соответствует ли оно выборке измеряемой области поведения.

Содержательная валидность – это один из основных типов валидности методики, характеризующей степень репрезентативности содержания заданий теста измеряемой области психических свойств. Такая валидность обычно применяется для тестов, исследующих деятельность, близкую или совпадающую с реальной, т.е. для тестов достижений, критериально-ориентированных тестов, а также методик, предназначенных для профотбора, анализа успешности овладения профессией и т.п.

Для валидизации личностных опросников и тестов интеллекта содержательная валидность используется ограниченно, лишь на начальных стадиях составления теста. Обусловлено это тем, что данные виды методик обычно не имеют того внутреннего

сходства с исследуемыми областями поведения, какое присуще тестам достижений, следовательно, анализ их содержания может разве что выявить гипотезы, приведшие составителя к выбору определенного материала для измерения тех или иных особенностей личности. Кроме того, личностные тесты, тесты интеллекта и способностей не опираются на конкретный курс обучения или на предшествующий жизненный опыт, исходя из которого отбираются задания теста. Поэтому способы выполнения одних и тех же заданий в тестах способностей и личностных опросниках, а также используемые ими психологические процессы будут отличаться друг от друга. Это означает, что один и тот же тест может измерять у разных людей разные функции. Например, студенты университета при решении задания теста на способности могут использовать вербальные формулировки, в то время как инженер будет решать ту же задачу с опорой на наглядные представления. Для установления валидности этих тестов необходима эмпирическая верификация гипотез [2].

При установлении содержательной валидности может создаться неоправданное впечатление о простоте ее проверки, заключающейся в просмотре содержания теста. Однако, по утверждению А. Анастази, решение не столь очевидно, как может показаться на первый взгляд [2]. Во-первых, необходим систематический анализ теста с целью охвата всего диагностируемого предмета, что предполагает подробное описание предметной области заранее, до составления теста. Во-вторых, правильно построенные тесты достижений должны охватывать цели обучения, а не его конкретные темы. В-третьих, следует избегать неоправданных обобщений в отношении области поведения, выборочно поверяемой тестом. Например, если орфографический тест измеряет способность распознавать правильно и неправильно написанные слова, то это не означает, что он также хорошо должен измерять правильность написания диктанта или другие аспекты орфографической грамотности [2].

Высокий уровень содержательной валидности обеспечивается уже на этапе подбора заданий. Обычно ему предшествует полная и систематическая процедура просмотра соответствующих учебников, учебных программ, а также анализ мнений специали-

стов. На основе полученной информации составляется спецификация теста для составителей заданий, которая включает охватываемые тестом области содержания, темы предмета, проверяемые учебные цели или способы действия, а также относительное значение каждой темы или способа действия. В заключение указывается требуемое число заданий каждого типа по каждой теме.

При обращении к специалистам необходимо указать их количество, профессиональный опыт, а также коэффициент согласованности их мнений при классификации или отборе заданий. Желательно также указывать и дату обращения к экспертам, что связано с постоянно меняющимися программами и содержанием учебных дисциплин, а также привести перечень проанализированных учебников и текстов программ с указанием года издания.

Помимо анализа предметной области содержательная валидизация может дополняться и рядом эмпирических методов. Например, отобранные задания можно подвергнуть соотнесению с успеваемостью. В этом случае в тесте оставляются те задания, которые дают прирост процента учащихся, переходящих с более низких уровней успеваемости на более высокие.

5.6.4. Другие виды валидности

Описанные выше виды валидности являются основными и должны использоваться при разработке или адаптации практически каждой психодиагностической методики. Однако наряду с ними существует ряд других, не столь распространенных, но все же достаточно существенных видов валидности. Это валидность ***синтетическая, очевидная и экологическая.***

Синтетическая валидность. Данный вид валидности является разновидностью валидности критериальной. Понятие синтетической валидности связано с понятиями отдельных и комплексных критериев. Чем более сложным, комплексным является изучаемый объект, тем шире и обобщенней должен быть критерий. На практике часто бывает трудно найти такой критерий, и поэтому оценка валидности производится при помощи разделения общего критерия на элементы (отдельные критерии) в соответствии с элементами оцениваемого качества или области поведения.

Очевидная валидность. Данный вид валидности отражает то, насколько «хорошим» тест кажется испытуемому или другому человеку, не располагающему сведениями о методике и целях ее использования. Очевидная валидность не относится к числу объективно определяемых видов валидности, поэтому подсчитать ее коэффициент невозможно. При этом, однако, она побуждает испытуемых к обследованию, способствует более серьезному и ответственному отношению к выполнению заданий теста и к заключениям, формулируемым психологом.

Экологическая валидность. Экологическую валидность можно определить как свойство теста, проявляющееся в том, что применение его для решения различных практических задач ведет к качественно различной интерпретации результатов тестирования. Экологическая валидность – это валидность теста по отношению к измеряемому свойству в контексте определенной ситуации [9].

5.6.5. Перекрестная валидизация

Необходимым условием для достижения высокой валидности теста является ее определение на выборке, отличной от той, на которой происходил отбор заданий. Эта независимая процедура определения валидности получила название перекрестной или кросс-валидизации. Это связано с тем, что при применении одной и той же выборки ее случайные свойства могут оказать влияние и привести к искусственному завышению коэффициента валидности. Хорошей иллюстрацией выше сказанного служит классическое исследование Кьюретона, проведенное в 1954 году [2].

В этом исследовании было показано, как при использовании одной и той же выборки для отбора заданий и проверки валидности теста можно получить фиктивный коэффициент валидности абсолютно при случайных условиях. В качестве критерия валидности автором были выбраны оценки 29 студентов, записавшихся на курс психологии. На первом этапе эксперимента, в соответствии с критерием, все студенты были поделены на две критериальные группы: оценки не ниже «B» (верхняя критериальная группа) и оценки ниже «B» (нижняя критериальная группа). Роль «заданий» в эксперименте выполняли номерки с числами от 1 до

85, написанными на одной стороне. Для получения тестового результата каждого из студентов номерки складывались в коробку, перемешивались и высыпались на стол. Те из них, которые выпадали цифрами кверху, регистрировались как номера, выполненных данным студентом заданий. Всего для каждого студента было выполнено 29 бросков указанного количества номерков, по результатам которых вычислялся совокупный показатель.

Следующим этапом эксперимента был анализ заданий с точки зрения выбранного критерия валидности. На этом основании из 85 «заданий» были отобраны те задания, которые чаще встречались в одной критериальной группе по сравнению с другой. Таких заданий оказалось 24, причем 9 заданий чаще встречались в верхней критериальной группе, а оставшиеся 15 заданий – в нижней критериальной группе. Первым 9 заданиям приписывался вес +1, а остальным заданиям приписывался вес –1.Затем с помощью данного ключа анализировались конкретные «ответы» студентов. Сумма весов заданий составляла суммарный тестовый балл каждого испытуемого.

На третьем этапе провели корреляционный анализ суммарного тестового балла с успеваемостью студентов. Несмотря на фиктивность теста и заведомо случайное происхождение тестовых баллов, корреляция успеваемости студентов с их суммарным баллом оказалось значимой и равной 0,82.

Данный пример иллюстрирует видимое соответствие между использованием одних и тех же случайных различий в отборе заданий и при определении валидности теста в целом. Опыт исследований показывает, что уменьшение валидности теста при перекрестной валидизации будет наибольшим, если: 1) выборки малы; 2) исходная совокупность заданий велика, а доля окончательно отобранных заданий мала; 3) если задания подбираются без заранее сформулированных гипотез, выводимых из теоретического и операционального определений.

Вопросы для самопроверки

1. В чем смысл понятия «валидность теста»?
2. Назовите основные виды валидности теста.

3. Какой вид валидности предполагает деление на текущую и прогностическую?

4. Что такое «контаминация» критерия?

5. Какой вид валидности можно оценить с помощью процедуры факторного анализа?

5.7. Стандартизация процедуры тестирования и обработки результатов

С целью снижения субъективно-психологической трудности теста, вызванной как особенностями тестирования, так и индивидуальными свойствами самого испытуемого, процедура и обработка результатов теста подвергается жесткой стандартизации [4]. Стандартизация касается объективных фактов процесса тестирования, куда входит тестовый материал, процедура проведения теста, обработка его результатов и условия материальной среды, а также субъективных факторов, включающих мотивацию и психические стояния испытуемого в момент проведения теста.

Стандартизация объективных факторов

В целях обеспечения объективности тестов их комплектация должна соответствовать определенным требованиям. В соответствии с международными стандартами в комплект любого теста входит три элемента: 1) руководство; 2) тестовая тетрадь; 3) регистрационные бланки и ключи для обработки результатов, а также нормы. В небольших по объему тестах нормы могут являться составной частью руководства [20].

Руководство к тесту пишется прежде всего для пользователя. В нем излагается назначение и психологическое содержание теста, дается описание его структуры, приводится подробнейшая процедура создания и проверки теста с указанием стандартизированных выборок и основных психометрических показателей и способов их вычисления. Кроме того, описывается процедура проведения теста, куда входит инструкция к тесту и его отдель-

ным субтестам, условия проведения, поведение экспериментатора.

Описание проведения теста включает в себя изложение условий проведения, материалов, требуемых при тестировании (карандаши, резинки, бланки ответа, кубики, картинки и т.п.), порядка их размещения в тесте и предъявления. Здесь же указываются временные ограничения, приводится полный текст сообщения для испытуемых, поведение экспериментатора.

Особое место уделяется правилам обработки и интерпретации результатов. Последняя должна включать ряд конкретных примеров, ее иллюстрирующих. В случае наличия нескольких форм предлагаемого теста указывается их специфика и назначение. Обязательным составным элементом руководства должны быть ограничения и противопоказания для применения теста [7].

Неотъемлемой составной частью предъявления теста испытуемым является инструкция, содержание которой во многом определяет объективность результатов тестирования.

Инструкция для испытуемого приводится в тестовой тетради или на регистрационном бланке, а также может быть включена и в руководство. Она включает в себя описание назначения теста, правил заполнения бланков ответов, а также демонстрационные примеры заданий.

Порядок составления инструкции включает в себя осуществление следующих мероприятий: 1) написание первого варианта инструкции в соответствии с целями теста; 2) пилотажное поведение теста на небольшой выборке (10 – 15 человек) с целью выяснения правильности ее понимания испытуемыми; 3) устранение явных ошибок в инструкции; 4) проведение инструкции психологами, не принимавшими участия в составлении первых ее вариантов на небольшой выборке испытуемых; 5) окончательная редакция инструкции [7]. Для того чтобы инструкция полностью отвечала целям тестирования, она должна быть составлена в соответствие с определенными правилами:

1. В инструкции должно быть сказано все существенное, что относится к выполнению теста. Упрощения, кажущиеся незначительными, могут существенно повлиять на результаты тестирования.

2. Инструкция должна быть краткой. Существует ошибочное мнение среди некоторых пользователей тестов, что повторение и дальнейшее объяснение достаточно ясных положений инструкции будет способствовать их лучшему пониманию. Напротив, повторение инструкции и снабжение ее дополнительными и подробными комментариями, с одной стороны, может ухудшить ее запоминание, а с другой стороны, создаст у испытуемого ложное впечатление о слишком большой трудности теста и невыполнимости его заданий.

3. При наличии в тесте разных по своему содержанию субтестов целесообразно, помимо общей инструкции, сформулировать и инструкции к каждому из них.

4. Инструкции должны быть однообразны и приспособлены к возрасту и уровню психического развития испытуемых. При этом инструкция должна быть ориентирована на «слабейшего».

5. Примеры заданий, иллюстрирующих понимание теста, должны быть тщательно подобраны и должны соответствовать задачам тестирования. Практика показывает, что наличие таких заданий позволяет снизить число не понявших инструкцию с 10% до 0,5%.

Вербальная форма подачи инструкции предполагает четкое, ясное и членораздельное ее предъявление с оптимальной степенью громкости.

Тестовая тетрадь предназначена для испытуемого и включает в себя набор заданий. В зависимости от характера теста она может представлять собой перечень утверждений или вопросов (личностные тесты-опросники) или набор вербальных и графических задач (тесты интеллекта и способностей). Задания в тесте располагаются, как правило, в порядке возрастания трудности. Исключение составляют личностные тесты-опросники и некоторые тесты скорости. Самые простые задания играют роль тренировочных, располагаются в начале теста (субтеста) и нумеруются 01, 02 и т.д. Обычно в тесте используются 1 – 2 примера. Назначение тренировочных заданий заключается в ознакомлении с принципами решения заданий и формировании первичных навыков работы с ним.

При конструировании тестовой тетради особое внимание необходимо обратить на графические тестовые задания, которые должны быть сделаны в соответствии с закономерностями восприятия и иметь определенную величину, яркость и четкость изображения.

Пользователю следует иметь в виду, что испытуемому категорически запрещается вносить любые пометки в тестовую тетрадь. Для этой цели ему предлагается лист ответов или регистрационный бланк.

Регистрационный бланк представляет собой специально организованный лист (или несколько листов) бумаги. Основная цель бланка – обеспечить удобство в заполнении и обработке результатов теста. В нем отражена последовательность заданий теста с вариантами ответа (в задачах с предписанными ответами) или с пропусками (в задачах со свободными ответами) [7]. В случае заданий с предписанными ответами варианты ответов могут задаваться в виде букв, цифр, которые испытуемый помечает заранее установленным способом (подчеркиванием, зачеркиванием, обводкой и т.п.).

Как правило, на бланке отводится место для фиксирования социально-демографической информации, даты и времени проведения тестирования, а также приводится шкала оценки. В случае индивидуального тестирования бланк (протокол) заполняется обычно экспериментатором, где могут указываться подсказки, временные ограничения или схемы вариантов решения.

Ключи к тесту представляют собой упорядоченные наборы правильных (а иногда и неправильных) ответов на вопросы теста или наборы ответов с оценками к вопросам для определения степени выраженности диагностируемой личностной характеристики. По форме ключи к тесту делят на стандартизированные и нестандартизированные [7].

Нестандартизированные ключи представляют собой напечатанные перечни правильных ответов или оценок по каждому варианту ответа. Такая форма оценка очень неудобна, поскольку занимает много времени.

Стандартизированные ключи применяются в виде шаблонов с отверстиями, которые по размеру точно совпадают с регистра-

ционным бланком. Правильные ответы (ответы в соответствии с ключом) обычно вырезаются и при совмещении с регистрационным бланком показывают степень точности ответа на данное задание.

Другим важным аспектом стандартизации объективных факторов является стандартизация процедуры и внешних условий тестирования.

Требования к внешним условиям тестирования обычно относятся к характеристикам помещения. Тестирование должно проводиться в хорошо проветриваемом помещении без световых, звуковых и прочих помех. Размеры помещения должны соответствовать установленным санитарно-гигиеническим нормам – не менее 2 кв. м на одного человека.

Процедура тестирования имеет свои особенности в зависимости от его типа. При индивидуальном проведении теста степень жесткости стандартной процедуры меньше, чем при групповом тестировании. Поэтому в качестве иллюстрации остановимся на требованиях к процедуре группового тестирования.

Непосредственная подготовка к проведению теста включает проведение следующих мероприятий: 1) проверка состояния помещения, его оснащения и пригодности для размещения определенного числа участников; 2) устранение ситуативных отвлекающих факторов; 3) проверка наличия состояния и размещения тестовых материалов.

Первое мероприятие предполагает оснащение помещения необходимым количеством столов и стульев. При этом следует иметь в виду, что каждый испытуемый должен располагаться за отдельным столом. Расстояние между столами должно быть достаточным для того, чтобы обеспечить удобный подход к каждому испытуемому, с одной стороны, и исключит возможность подсказок и подглядывания, с другой стороны. Желательно, чтобы экспериментатор находился на возвышении для более удобного обзора ситуации тестирования.

Освещенность рабочего места должна быть равномерной при интенсивности не ниже 500 люксов.

Устранение отвлекающих ситуационных факторов предполагает устранение а) различного рода шумов (уличных, в данном

здании, радио- и телеаппаратуры, гудения ламп дневного света и т.п.), а также других акустических помех; б) запахов; в) световых помех (мигание света); г) неопрятности помещения.

Для обеспечения непрерывной работы на каждом рабочем столе должны лежать: тестовая тетрадь и регистрационный бланк без всяких посторонних записей и пометок, остро отточенный карандаш и ластик.

Следует также учитывать ряд факторов, влияющих на результаты тестирования: 1) время проведения теста. Наиболее благоприятное время – утренние часы, начиная с 9.00; 2) наличие отработанной процедуры тестирования; 3) наличие делового дружеского контакта между участниками тестирования [7, 20].

Процедура проведения теста должна проходить в соответствии со стандартными требованиями и полностью соответствовать процедуре, изложенной в руководстве. Она включает в себя следующие основные этапы[8]:

1. Объяснение целей и задач тестирования, гарантирование соблюдения профессиональной тайны и конфиденциальности процедуры.

2. Изложение инструкции испытуемым с предоставлением им возможности потренироваться в решении примеров. Проверка правильности понимания инструкции. Сообщение временных ресурсов, правил исправления ошибок, правил поведения во время тестирования.

3. Предложение к испытуемым заполнить паспортные и биографические данные в регистрационном бланке и проверка правильности его заполнения.

4. После ответов на вопросы испытуемых подача команды начала выполнения теста. На этом этапе важно следить за работой испытуемых, акцентируя внимание на строгом выполнении всех команд экспериментатора, а также за дисциплиной во время выполнения заданий.

5. По истечении времени выполнения теста подача команды об окончании работы. При этом все материалы тестирования

[8] Более подробная процедура проведения изложена в работе В. Гайды и В. Захарова [7].

должны быть сложены в определенном порядке на краю стола, для того чтобы экспериментатор или его помощник мог их собрать. На этом этапе также очень важным моментом является единообразное выполнение всеми испытуемыми команд экспериментатора, что позволяет создать равные условия выполнения теста.

6. В заключение производится подсчет всех бланков ответов на предмет соответствия количеству испытуемых, проверка правильности заполнения.

Стандартизация субъективных факторов

Стандартизация процедуры и условий тестирования хотя и требует от экспериментатора определенных усилий, однако не представляет большой трудности. Гораздо сложнее нивелировать влияние субъективных факторов, где главную роль играет мотивация испытуемых и специфика их психических состояний.

Мотивы, которые побуждают испытуемых принять участие в тестировании, могут быть абсолютно разными, что существенным образом отразится на результатах тестирования. Первые попытки стандартизации мотивации испытуемых достигаются на этапе составления заданий. В этом случае необходимо учитывать следующие требования к формулировке заданий теста:

1. Задания теста должны быть легко и однозначно понимаемы испытуемыми.

2. Они должны быть нетривиальными и возбуждать у испытуемых интерес к тесту.

3. Задания должны быть современными и отражать реалии жизни испытуемых. Например, если вы в качестве задания математического субтеста воспользуетесь задачами из учебника арифметики позапрошлого века о купце, поехавшем на рынок и купившем 6 пудов зерна, 8 аршин ситца, 3 фунта ситного и т.п., то вряд ли это вызовет большой интерес к решению теста у сегодняшних испытуемых.

Следующим этапом нивелирования мотивации является начало процедуры тестирования, где экспериментатор должен четко и ясно изложить цели тестирования, объяснить причины, по которым испытуемые должны приложить максимум усилий для выполнения теста. При этом он может либо акцентировать вни-

мание на возможности испытуемых проверить свои способности или актуализировать соревновательный мотив, если проводится групповое тестирование. Важным моментом в этом плане является и гарантия со стороны экспериментатора сохранения тайны индивидуальных ответов.

Достаточно сложным моментом при достижении равных субъективных условий при тестировании является уравнивание испытуемых с точки зрения их психических состояний. Поскольку учесть состояние каждого испытуемого довольно трудно и готовых рецептов в этом плане не существует, ряд авторов предлагают перечень категорий тех лиц, которых нежелательно допускать к тестированию [7]:

1. К тестированию не допускаются лица, находящиеся в любой степени алкогольного или наркотического опьянения или под действием других фармакологических средств.
2. К тестированию не допускаются соматически больные люди или лица, жалующиеся на недомогание.
3. К тестированию не допускаются лица, находящиеся в предстрессовом или послестрессовом состоянии (возбуждении, депрессии и т.д.).
4. К тестированию не допускаются лица после тяжелой или утомительной работы или жалующиеся на усталость.
5. К тестированию не допускаются лица, негативно настроенные по отношению к тестовым испытаниям или решительно отказывающиеся от них.

Указанные мероприятия позволят обеспечить равные условия для всех испытуемых, выполняющих данный тест, и снизить влияние случайных факторов, влияющих на результаты тестирования.

Вопросы для самопроверки

1. Какие компоненты должен включать в себя любой тест?
2. Какие требования предъявляются к инструкции к тесту?
3. Зачем нужны тренировочные задания в тесте?
4. Какой контингент лиц не допускается к тестированию?

Глава 6. **Интерпретация тестовых показателей**

Большинство тестов дают непосредственные числовые сведения об исполнении человеком теста. Такую оценку принято называть в психометрике ***«сырым» баллом***. Она может выражаться в количестве правильно решенных задач, времени выполнения задания или подобных сведениях. В связи с легкостью ее получения многие пользователи начинают интерпретировать результаты тестирования, не осознавая ограниченности такой оценки. Взятая сама по себе, «сырая» оценка по психологическому тесту не имеет значимости. Ее интерпретация возможна только путем сопоставления с каким-либо стандартом.

Данное положение вытекает из самой природы психологического измерения, интерпретация которого иная, нежели измерения физического. Большинство шкал физических измерений – это шкалы отношений, которые имеют абсолютную нулевую точку и равные расстояния между делениями шкалы (метрические шкалы длины, веса и т.п.).

Это позволяет сказать, что один мальчик вдвое длиннее другого или что он достиг 60% своего вероятностного взрослого роста. О психологических измерениях таких утверждений делать нельзя, в связи с тем, что «сырая» оценка не отражает реальную дистанцию между индивидами. Например, по одному тесту Петя решил 18 задач, а Вася – 16 задач, а по второму тесту Петя решил 6 задач, а Вася – 4. Можно ли сказать, что Петя умнее Васи на 2 балла, что 2 балла в первом и втором случае – эквивалентны? Нет, потому что разница в два балла может быть не значимой и соответствовать одному и тому же уровню интеллекта. Использование в качестве относительной оценки процентного соотношения также не является надежным, поскольку разный процент решения задач по трем разным субтестам может отражать абсолютно одинаковое

их количество (30% и 60% решенных задач по двум субтестам соответствует 30 решенным задачам в обоих случаях).

Приведенные примеры показывают, что первичные результаты любого психологического теста остаются бессмысленными без дополнительных данных. На практике они интерпретируются путем сопоставления с нормами выполнения данного теста в выборке стандартизации или другими стандартными показателями. Существует три вида интерпретации: нормативно-референтная, содержательно-референтная и критериально-референтная [20].

6.1. Нормативно-референтная интерпретация

Данный вид интерпретации тестовых баллов является традиционным и наиболее распространенным в практике психологического тестирования. ***Норма представляет собой количественную меру оценки результата тестирования, определяющую степень выраженности психологического свойства как объекта измерения, которая показывает, какое место занимает данный индивид в эталонной группе людей с определенными признаками (социально-демографическими, половыми, возрастными и т.п.).***

Тестовые нормы как некие относительные (производные) показатели служат двум целям: 1) они указывают положение обследуемого относительно эталонной (нормативной) выборки, что позволяет оценить выполнение им теста на фоне других лиц; 2) нормы позволяют непосредственно сравнивать данные, полученные с помощью разных тестов.

Любая норма, в чем бы она ни выражалась, ограничивается конкретной группой людей, для которой она вырабатывалась. Это означает, что нормы в психологическом тестировании не могут быть абсолютными, постоянными и универсальными. Они отражают выполнение теста испытуемым из определенной выборки стандартизации. При формировании такой выборки пытаются получить репрезентативный срез популяции, на которую ориентирован тест. В статистике обычно используют термины: «генеральная совокупность» («популяция», «генеральная популяция»

ляция») и «выборка» («выборочная совокупность», «выборочная популяция») [2], [4], [7], [17], [33].

Генеральная совокупность – множество элементов, объединенных общей характеристикой, указывающей на их принадлежность к определенной системе [5], [34]. Примером генеральной совокупности может служить все взрослое русскоязычное население нашей страны.

Однако на практике охватить тестированием всю генеральную совокупность невозможно, поэтому из этой совокупности извлекается некая ее репрезентативная (представительная) часть, получившая название выборка.

Выборка – это часть элементов генеральной совокупности, являющаяся репрезентативной по отношению к ней, но имеющая меньший объем [4]. Качественный и количественный состав выборки определяется характером и природой изучаемого явления и определяется его универсальностью и однородностью генеральной совокупности [7].

При разработке тестовых норм на выборку стандартизации следует обращать существенное внимание. Для обеспечения стабильности нормативных значений выборка должна быть достаточно большой по объему, чтобы при отборе другой аналогичной выборки из той же популяции не было существенных различий в нормативных данных. Объем выборки может варьироваться в широких пределах, но на практике должен приближаться к минимальному порогу, необходимому для получения достоверных результатов. В статистике существуют довольно точные способы определения объема выборки [4], [7], однако для простого уменьшения ошибки нормативная выборка в количестве 200 человек и более подходит при соблюдении принципов случайности и пропорциональности отбора испытуемых.

Размер выборки стандартизации является важной, но далеко не основной ее характеристикой. Другой особенностью выборки является её качественный состав. Она должна охватывать все группы генеральной совокупности по основным социально-демографическим характеристикам. Например, процентное соотношение мужчин и женщин в выборке стандартизации должно соответствовать процентному соотношению данного признака в ге-

неральной совокупности. С этой целью выборка подвергается стратификации по различным переменным, среди которых ведущими являются: пол, возраст, социальный статус, уровень образования и т.д. Для каждой стратификационной переменной обычно берется выборка в количестве 300 человек. При наличии нескольких стратификационных переменных объемы выборки могут быть достаточно большими. Поэтому на практике подавляющее большинство тестов стандартизировано не для больших популяций, и многие создатели тестов делают специфические нормы для групп, релевантных природе и практическому применению теста.

Существуют различные классификации тестовых норм. Можно выделить универсальные нормы, групповые нормы, локальные нормы.

Универсальные нормы устанавливаются для широкого контингента лиц. Они в малой степени зависят от таких признаков, как пол, возраст и т.п. Например, нормы для лиц в возрасте от 16 до 60 лет. Различия между группами в соответствии с этими признаками могут быть статистически незначимыми.

Групповые нормы разрабатываются для больших групп людей, отличающихся наличием какого-либо признака. Примерами таких норм могут быть профессиональные нормы, национальные нормы, возрастные нормы и т.д.

Локальные нормы устанавливаются и применяются для узких категорий людей в соответствии с конкретными целями тестирования. Например, нормы для отбора претендентов на соответствующую работу будут включать возраст, пол, уровень и специфику образования (коэффициент интеллекта для мужчин в возрасте от 25 до 30 лет, имеющих высшее гуманитарное образование).

Процедура разработки норм связана с использованием закона нормального распределения. Поэтому на практике, прежде чем перейти к построению норм, необходимо проверить предположение о нормальном распределении полученных тестовых данных. Эта проверка осуществляется путем сопоставления сведений о свойствах полученных величин и условий изучения со свойствами функций нормального распределения. Данное сопоставление производится сначала с помощью качественных мето-

дов, а потом подкрепляется специальной количественной обработкой.

Для качественного анализа характера распределения осуществляется графическое представление полученных результатов тестирования путем составления таблицы частотного распределения и построения графика. График может быть построен в виде полигона частот или гистограммы.

Графическое представление позволяет провести первичный качественный анализ на предмет выявления наличия или отсутствия асимметрии кривой (левосторонней или правосторонней), ее двугорбости или иных аномалий.

Количественное сопоставление включает в себя ряд этапов.

Первый этап предполагает сравнение отдельных свойств эмпирического распределения со свойствами нормального закона. Конкретизируется это в уточнении результатов качественного анализа и вычислении коэффициентов асимметрии и эксцесса.

Асимметрия выражается графически в виде скошенной вариационной кривой, вершина которой может быть сдвинута от центра распределения либо влево (левосторонняя асимметрия), либо вправо (правосторонняя) асимметрия. Определяется она по формуле:

$$A_s = \frac{1}{s_x^3}\left(\Theta - 3C_x^2 + 2\bar{x}^3\right),$$

где $\bar{x}$ – среднее арифметическое, s_x – стандартное отклонение, Θ – среднее кубическое $\left(\Theta = \sqrt[3]{\frac{1}{n}\sum x^2}\right)$, С – среднее квадратическое отклонение $\left(C = \sqrt{\frac{1}{n}\sum x^2}\right)$.

Графическое выражение эксцесса демонстрируют плоско- и островершинные кривые, а также двух- и многовершинные эмпирические кривые. Эксцесс кривой (E_x) определяется по формуле:

$$E_x = \frac{1}{s_x^4}\left(Q^4 - 4\Theta^3\bar{x} + 6C^2\bar{x}^2 - 3\bar{x}^4\right) - 3,$$

где Q – среднее значение четвертой степени $\left(Q = \sqrt[4]{\frac{1}{n}\sum x^4} \right)$.

Показатели асимметрии и эксцесса нормального распределения равны нулю. При отклонении от данного значения можно говорить об анормальности распределения. Допустимые значения коэффициентов асимметрии и эксцесса устанавливаются с помощью статистических таблиц.

Второй этап проверки нормальности распределения заключается в построении теоретической функции распределения по эмпирическому ряду, предполагающему сравнение эмпирической и теоретической вероятностей распределения с помощью критериев Пирсона или Колмогорова и Смирнова [4], [7].

Дальнейшая процедура разработки тестовых норм связана с выбором типов шкал, с помощью которых будет производиться сравнение результатов теста с групповыми данными. Одной из распространенных в психологической диагностике форм такого оценивания является расчет *процентилей (перцентилей).*

Процентиль представляет собой процентильный ранг, который соответствует проценту лиц в выборке стандартизации, получивших равный или более низкий балл, чем балл данного испытуемого. Например, если 30 процентов лиц решают 15 задач в тесте на арифметические мышление, то первичному показателю 15 будет соответствовать 30 процентиль (P_{30}). Процентили указывают на относительное положение испытуемого в выборке стандартизации и представляют собой специфическую ранговую шкалу, которую можно назвать «ранговая шкала наоборот». Такое название предполагает ранжирование не сверху от лучшего члена группы (1 ранг) до члена группы, находящегося на 100 месте (100 ранг), а снизу, т.е. чем ниже процентиль, тем хуже позиция индивида в группе.

Процентили имеют ряд преимуществ перед остальными шкалами стандартизации: их легко рассчитать и понять даже неподготовленному человеку; их интерпретация не требует наличия нормального распределения данных тестирования; они универсальны, т.е. легко применимы для любых выборок и типов тестов.

Вместе с тем данный способ оценивания результатов не свободен от недостатков. Поскольку процентили являются порядковыми баллами, то они не годятся для использования более мощной параметрической статистики. Одним из основных недостатков процентильной шкалы является то, что арифметически одинаковые различия процентильных тестовых оценок могут не соответствовать равным различиям в выраженности оцениваемого свойства. Например, нельзя утверждать что испытуемые А и В, получившие соответственно процентильные оценки 10 и 20, имеют те же различия, что и испытуемые С и Д, получившие процентильные оценки 40 и 50. В случае нормального распределения первое различие будет превышать второе, поскольку переменные группируются более тесно в центре распределения и рассеиваются по мере удаления к его краям. Процентили показывают относительное положение каждого испытуемого в нормативной выборке, но не величину различий между результатами, что создает некоторое неудобство в интерпретации данных.

Более широкое распространение в психодиагностике получили стандартные показатели, основанные на использовании *интервальной шкалы.*

Основная особенность интервальной шкалы заключается в наличии равенства единиц измерения показателя теста на всем интервале его изменений. Это выражается в отклонении индивидуального результата от средней нормы в единицах, пропорциональных стандартному отклонению распределения. Такое преобразование получило название линейного, так как предполагает вычитание из каждого первичного показателя одной и той же величины с последующим делением результата на другую постоянную величину. Линейно преобразованные стандартные показатели, получившие в психометрике название *Z -оценки*, рассчитываются по формуле:

$$Z_i = \frac{X_i - \overline{X}}{\sigma},$$

где $\overline{X}$ – среднее арифметическое, σ – стандартное отклонение.

Z-оценки не всегда удобны для практической работы, т.к. они могут быть дробными, положительными и отрицательными. Однако их можно легко перевести в любую другую более удобную шкалу, которая представляет собой стандартную шкалу с заданным средним значением и средним квадратичным отклонением. Такое преобразование осуществляется с помощью следующей формулы:

$$S_c = \frac{X - \overline{X}}{\sigma} \cdot A + M,$$

где A – заданное среднее квадратичное отклонение, M – заданное среднее значение.

В настоящее время в психодиагностике используются следующие виды стандартных психометрических шкал:

1. Шкала IQ Векслера: IQ = 100 + 15z.
2. Шкала IQ Амтхауэра: IQ = 100 + 10z.
3. Шкала Т Маккола: T= 50 + 10z.
4. Шкала «стэнайнов» (стандартная девятка) Д. Гилфорда: C= 5+2z.
5. Шкала «стэнов» (стандартная десятка) Р. Кеттелла: C = 5,5 + 2z.
6. Шкала школьных оценок: S = 3+1z.
7. Шкала методики SAT (CEEB) для оценки способности к обучению: S = 500 + 100z.

Среди наиболее часто применяемых шкал стандартных показателей необходимо отметить шкалы IQ Д. Векслера и Р. Амтхауэра, которые используются при расчете нормативных показателей для тестов интеллекта, тестов способностей и тестов достижения, а также шкалу «стэнов» Р. Кеттелла, используемую при расчете нормативных показателей для личностных опросников.

На рис. 4 приводятся соотношения основных шкал стандартных показателей при нормальном распределении переменной.

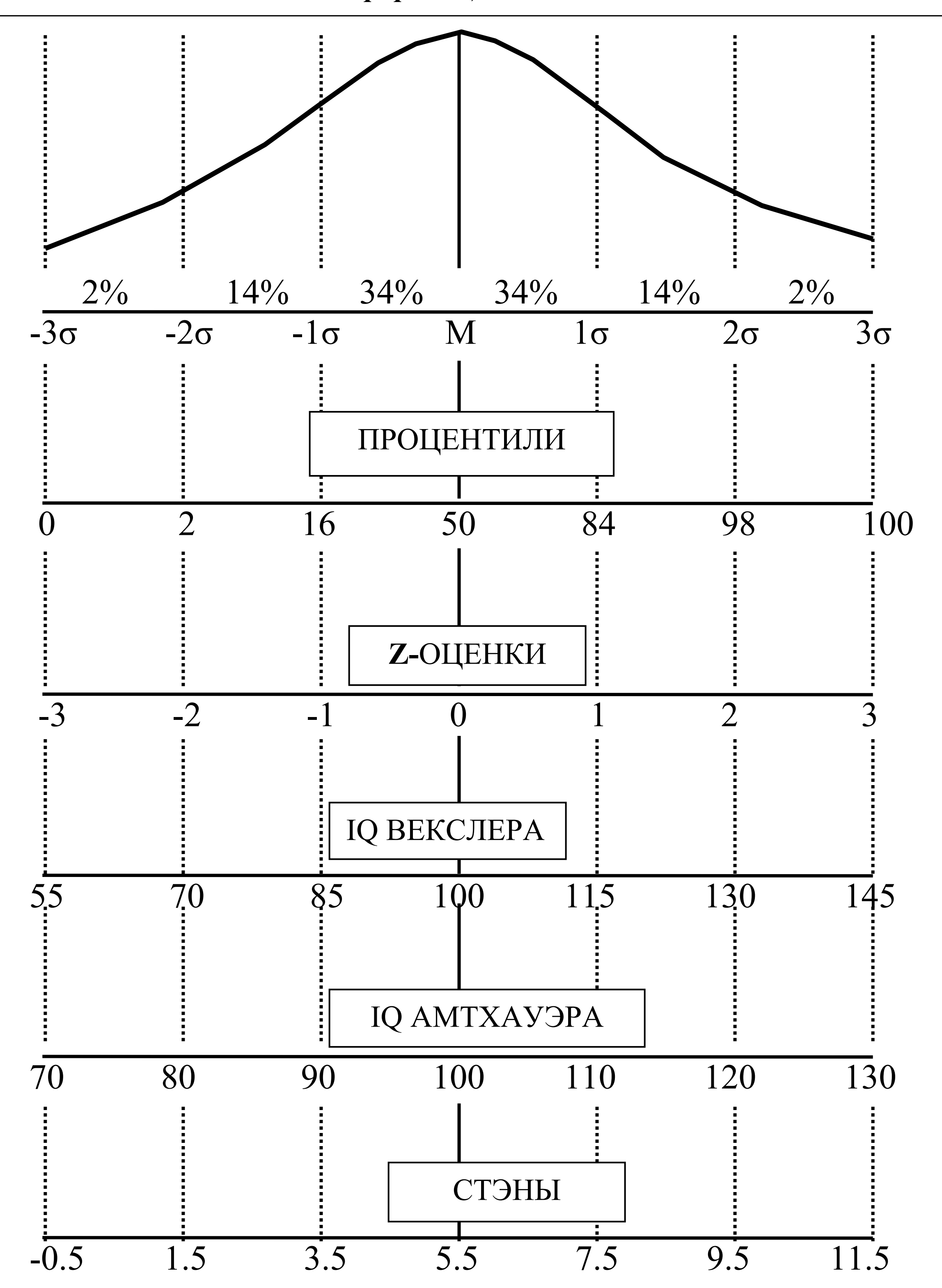


Рис. 4. Соотношение основных шкал стандартных показателей

Одной из причин введения производной шкалы вместо первичных показателей является стремление к сопоставимости показателей различных тестов. Но для того, чтобы результаты были

сопоставимы, они должны иметь одинаковое исходной распределение, а именно – нормальное. Если эмпирическое распределение отличается от нормального, оно может быть в большинстве случаев искусственно нормализовано путем нелинейного преобразования первичных оценок во вторичные нормально распределенные оценки. Такой метод называется «нормализация по составу» [34]. Под нормализованными стандартными показателями понимают стандартные показатели, соответствующие распределению, преобразованному так, что оно принимает вид нормальной кривой. Одним из способов нормализации по составу является перевод исходных оценок в процентильные (таблица 2).

Таблица 2

Процедура проведения нормализации по составу

X	P	F	F^*	PR	Z
3	2	2	1	1,6	– 2,1
4	18	20	11	17,7	– 0,9
5	13	33	26,5	42,7	– 0,2
6	8	41	37	59,7	0,2
7	10	51	46	74,2	0,6
8	6	57	54	87,1	1,1
9	4	61	59	95,2	1,7
10	1	62	61,5	99,2	2,4
		N=62		Σ = 100	M=0 σ = 1

В столбце X содержатся сырые баллы по тесту (не нормализованные). В столбце P – частоты встречаемости сырых баллов в выборке из 62 испытуемых. В столбце F – кумулятивные (т.е. накопленные от строчки к строчке частоты). Данные столбца F^* (кумулятивные баллы) вычисляются на основании данных столбцов 2 и 3 по формуле: $F_i^* = F_i - \frac{1}{2}P_i$. В столбце PR представлены процентильные ранги $(PR_i = F^* \times 100/\ n)$. В столбце Z даются нормализованные баллы (Z-оценки), полученные из соответст-

вующих процентильных рангов с помощью статистических таблиц.

Таким образом, алгоритм по анализу распределения тестовых баллов и построения тестовых норм выглядит следующим образом:

1. Формирование выборки стандартизации (случайной и стратифицированной по какому-либо параметру) из той популяции, на которой предполагается применение теста и проведение теста.
2. Группировка «сырых» баллов с учетом выбранного интервала квантования.
3. Построение распределения частот тестовых оценок в виде таблицы или соответствующих графиков и гистограмм.
4. Расчет среднего значения, стандартного отклонения, а также коэффициентов асимметрии и эксцесса.
5. Проверка нормальности распределения.
6. При подтверждении гипотезы о нормальности распределения тестовых показателей проведение линейной стандартизации с расчетом стандартных оценок.
7. В случае небольшого отличия эмпирического распределения от нормального и при наличии теоретических оснований для ожидания получения такового, проведение нормализации по составу.
8. В случае сильных расхождений эмпирического распределения и нормального, обследование более широкой выборки испытуемых или отказ от использования данного теста.

Рассматривая вопросы интерпретации тестовых показателей и сравнения их с нормами, следует учитывать относительность нормативных баллов.

Относительность норм проявляется, прежде всего, в разных местах, занимаемых человеком в различных выборках стандартизации. Например, испытуемый, имеющий высокий процентильный балл по шкале технических интересов в возрастной выборке стандартизации, может быть эквивалентен низкому баллу при сравнении его с профессиональными инженерно-техническими работниками [4], [20].

С другой стороны, нормативные данные могут изменяться в связи с изменениями социальных, экономических и других условий жизни человека. В частности, изменения условий жизни в нашей стране за последние 15 лет привели к трансформации ценностных ориентаций, интересов людей, что в свою очередь потребовало корректировки нормативных оценок в соответствующих тестах. Поэтому нормы тестирования должны подвергаться проверке и корректировке в случае существенных изменений, происходящих в обществе и влияющих на поведение человека и его личностные особенности.

Помимо сравнения показателей с групповыми нормами может иметь место сравнение тестовых оценок испытуемого с нормой индивидуальной. Такие оценки в психометрике получили название *ипсативных оценок*.

Ипсативные оценки нацелены на определение внутрииндивидуальных соотношений и не затрагивают межиндивидуальных различий [16]. Например, мерой мотивации к какому-либо виду деятельности могут быть временные и финансовые затраты. Покупка музыкального инструмента (флейты) за довольно значительную сумму двумя людьми, с точки зрения нормативного подхода, означает равенство их мотивации, что может быть само по себе ошибкой, поскольку не учитываются индивидуальные различия испытуемых (разный уровень способностей, разный доход). Одинаковые траты богатого человека и студента консерватории в этом случае будут свидетельствовать о большей степени мотивации последнего. Во избежание такого рода ошибок показатель ипсатируется, т.е. мерой мотивации будет отклонение от среднего значения денежных затрат данного индивида, но в других видах деятельности [13].

6.2. Критериально-референтная интерпретация

В отличие от рассмотренного выше нормативного измерения, *критериальное измерение* основано на том, что может делать и что знает испытуемый, а не как он выглядит на фоне других лиц. Оно направлено на оценку компетентности испытуемого в четко

определенной области, а не на измерение абстрактных свойств. В этой связи критериальное измерение используется в тестах достижений, способностей. Обычно все испытуемые должны достичь определенного уровня компетентности безотносительно друг друга. Оценка результата заключается не в определении места индивида на континууме по степени выраженности какой-либо черты, а в достижении некоего критерия, установленного эмпирическим или экспертным путем. Испытуемые в этом случае оцениваются по принципу «все или ничего», достиг определенного уровня или нет. Например, для продолжения образования в американских университетах претенденты из числа российских студентов должны сдать тест на знание английского языка и набрать определенное количество баллов. Студенты, не набравшие требуемого количества баллов, исключаются из дальнейшего конкурса.

6.3. Содержательно-референтная интерпретация

Содержательно-референтная интерпретация имеет место при непосредственной интерпретации балла в соответствии с деятельностью на каждой точке измеренного континуума достижений [21].

Вопросы для самопроверки

1. Какой статистический закон лежит в основе процедуры нормирования?
2. В чем состоят основные цели нормирования тестовых показателей?
3. Что такое «шкалы стандартных показателей»?
4. Как соотносятся между собой понятия « генеральная совокупность» и «выборка»?
5. Приведите пример порядковых шкал стандартных показателей?
6. Как называется шкала, имеющая среднее значение М= 5,5 и стандартное отклонение $\sigma = 2$?
7. Как определить место испытуемого в выборке стандартизации с помощью стандартных шкал?

Заключение

В настоящей работе была предпринята попытка рассмотреть основные положения теории и практики создания и применения психодиагностических методик. Были изложены вопросы, касающиеся предмета психодиагностики и основных теоретических концепций, лежащих в основе его определения, раскрыты основные этапы процедуры создания и проверки теста. При этом мы не случайно ограничились только классической эмпирико-статистической теорией создания теста.

Во-первых, данный подход является традиционным в психодиагностике и его описание содержится в многочисленных пособиях, учебниках и монографиях.

Во-вторых, на основе эмпирико-статистической теории создано подавляющее большинство тестовых методик.

Естественно, что не все вопросы классического подхода в психодиагностике получили подробное освещение в силу ограниченного объема учебного пособия. Мы остановились на наиболее принципиальных, на наш взгляд, моментах теории тестов и процедуры их создания. За пределами нашего анализа остались некоторые вопросы, связанные с начальными этапами создания тестовых методик, а также этические принципы создания и использования тестовых методик. Кроме того, мы не стали давать традиционный для многих учебников обзор основных тестовых методик.

Анализ современного состояния психодиагностики показывает, что традиционная процедура создания тестовых методик и лежащая в ее основе теория личностных диспозиций имеют существенные ограничения, что дало повод некоторым исследователям говорить о «кризисе в психодиагностике» [65]. На наш взгляд, речь, по-видимому, идет не о кризисе в психодиагностике, а о появлении ряда проблем, решение которых позволит обеспечить дальнейшее ее развитие.

В теоретическом аспекте плодотворной является концепция лично-ситуационного взаимодействия, практическое применение которой позволит разрабатывать принципиально новые методики, учитывающие не только выраженность тех или иных личностных черт, но и влияние ситуации.

Другая важная особенность современной психодиагностики – появление ряда новых подходов к конструированию тестов, среди которых достаточно популярной становится стохастическая теория тестов – Item Response Theory (IRT). Она опирается на концепции латентно-структурного анализа, предложенного П. Лазарсфельдом и являющегося разновидностью многомерного анализа данных. Главной отличительной особенностью этой теории по сравнению с классической моделью является отсутствие фундаментальных понятий традиционной теории, а именно валидности и надежности. Указанная концепция реализуется рядом статистических моделей, среди которых наиболее известны модель Г. Раша и модели А. Бирнбаума [2], [9], [12].

Таким образом, разработка новых направлений в теории и практике применения психодиагностических методик будет способствовать не только совершенствованию теоретических концепций и процедуры создания тестов, но и даст толчок для развития традиционной психодиагностики.

Литература

1. Аванесов, В.С. Тесты в социологическом исследовании / В.С. Аванесов.– М.: Наука, 1982.

2. Анастази, А. Психологическое тестирование. 7-издание / А. Анастази, С. Урбина. – СПб.: Питер, 2001.

3. Бурлачук, Л.Ф. Психодиагностика: учебник для вузов / Л.Ф. Бурлачук. – СПб.: Питер, 2003. – 352 с.

4. Бурлачук, Л.Ф. Словарь-справочник по психодиагностике / Л.Ф. Бурлачук, С.М. Морозов. – СПб.: Питер, 1999.

5. Бурлачук, Л.Ф. Психология жизненных ситуаций: учеб. пособие / Л.Ф. Бурлачук, Е.Ю. Коржова. – М.: Российское Педагогическое Агентство, 1998.

6. Витцлак, Г. Основы психодиагностики / Г. Витцлак // Психодиагностика: Теория и практика. Пер. с нем. / под ред. Н.Ф. Талызиной. – М.: Прогресс, 1986.

7. Гайда, В.К. Психологическое тестирование / В.К. Гайда, В.П. Захаров. – Л.: ЛГУ, 1982.

8. Дружинин, В.Н. Психологическая диагностика способностей: теоретические основы: в 2 ч. / В.Н. Дружинин. – Саратов: Изд-во СГУ, 1990.

9. Дружинин, В.Н. Экспериментальная психология: учеб. пособие / В.Н. Дружинин. – М.: ИНФРА-М, 1997.

10. Иванова, Н.Л. Качественный анализ: репертуарные личностные методики: учеб. пособие / Н.Л. Иванова. – Ярославль: ЯГПУ, 1999.

11. Келли, Дж. Психология личности. Теория личностных конструктов / Дж. Келли. – СПб.: Речь, 2000.

12. Клайн, П. Справочное руководство по конструированию тестов / П. Клайн. – Киев: Пан Лтд., 1994.

13. Козлова, И.Н. Личность как система конструктов / И.Н. Козлова // Системные исследования. Ежегодник. – 1975. – М.: Наука, 1976.

14. Конева, Е.В. Судебно-психологическая экспертиза: учеб. пособие / Е.В. Конева, В.Е. Орел / Ярославль: Яросл. гос. ун-т, 1998.

15. Кодекс этических принципов психолога // Психологический журнал, 1987, № 5.

16. Кулагин, Б.В. Основы профессиональной психодиагностики / Б.В. Кулагин.– Л.: Медицина, 1984.

17. Лакин, Г.Ф. Биометрия: учеб. пособие. – 4-е изд. / Г.Ф. Лакин. – М.: Высш.шк., 1990.

18. Личностная анкета KUD (руководство к проведению и интерпретация): метод. указания / сост. В.Е. Орёл, В.И. Чирков. – Ярославль, ЯрГУ, 1988.

19. Мельников, В.М. Введение в экспериментальную психологию личности / В.М. Мельников, Л.Т. Ямпольский. – М.: Просвещение, 1985.

20. Нормы профессиональной этики для разработчиков и пользователей психодиагностических методик. Стандартные требования к психологическим тестам. 2-е изд. / сост. А.А. Рукавишников, И.Г. Сенин, М.В. Соколова, А.П. Чернявская. – Ярославль: НПЦ «Психодиагностика», 1998.

21. Общая психодиагностика / под ред. А.А. Бодалева и В.В. Столина. – М.: Изд-во МГУ, 1987.

22. Опросник профессиональных предпочтений // А.Н. Воробьев, И.Г. Сенин, В.И. Чирков (авторы адаптации): руководство. 2-е изд. – Ярославль, НПЦ «Психодиагностика», 1997.

23. Орёл, В.Е. Опыт адаптации методики диагностики личностных характеристик NEO PI R / В.Е. Орёл, И.Г. Сенин, Т.А. Мартин // Вестник клинической психологии. – 2003. – Т. 1. – № 2. – С. 193-199.

24. Первин, Л. Психология личности: теория и исследования / пер. с англ. М.С. Жамкочьян под ред. В.С. Магуна / Л. Первин, О. Джон. – М.: Аспект Пресс, 2000.

25. Петренко, В.Ф. Психосемантика сознания / В.Ф. Петренко.– М.: Изд-во МГУ, 1988.

26. Петровский, А.В. История советской психологии / А.В. Петровский. – М.: Политиздат, 1967.

27. Платонов, К.К. Проблема способностей / К.К. Платонов. – М.: Наука, 1972.

28. Попова, Л.А. Апробация теста MBTI: курсовая работа / научн. руководитель А.А. Рукавишников / Л.А. Попова. – Ярославль, 1997.

29. Пособие по применению личностного опросника Р. Кеттелла 16PF / сост. А.А. Рукавишников, М.В. Соколова. 3-е изд., Ярославль, НПЦ «Психодиагностика», 2000.

30. Похилько, В.И. Психодиагностика индивидуального сознания / В.И. Похилько // Общая психодиагностика / ред. А.А. Бодалев, В.В. Столин.– М.: Изд-во МГУ, 1987.

31. Проблемы профконсультирования молодежи // Материалы республикан. научно-практич. конф. – Таллинн, 1989.

32. Профессиональный кодекс этики для психолога (ФРГ) // Вопросы психологии. – 1990. – № 6.

33. Процесс социального исследования // пер. с нем. / общ. ред. Ю.Е. Волкова. – М.: Прогресс, 1975.

34. Психодиагностика: теория и практика. Пер. с нем. / под ред.Н.Ф. Талызиной.– М.: Прогресс, 1986.

35. Психологическая диагностика: учебное пособие / под ред. К.М. Гуревича и Е.М. Борисовой.– М.: Изд-во УРАО, 1997.

36. Психологическая наука в России XX века. Проблемы теории и истории / под ред. А.В. Брушлинского. – М.,1997.

37. Психолого-педагогическая профконсультация учащихся: метод. рекомендации / под общ. ред. Е.А. Климова. – Л.: ВНИИпрофтехобразов., 1975.

38. Психодиагностические методы (в комплексном лонгитюдинальном исследовании студентов) // под ред. Б.Г. Ананьева.– Л.: Изд-во Ленингр. ун-та, 1976.

39. Родионов, А.В. Психодиагностика спортивных способностей / А.В. Родионов. – М.: Физкультура и спорт, 1973.

40. Росс, Л. Человек и ситуация. Перспективы социальной психологии / Л. Росс, Р. Несбит. – М.: Аспект Пресс, 1999.

41. Руководство к личностному опроснику НЭП / А.А. Рукавишников Л.А. Соснина (авторы адаптации). – Ярославль, НПЦ «Психодиагностика», 1999.

42. Сидоренко, Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: ООО «Речь», 2002.

43. Тест умственных способностей / И.Г. Сенин, О.В. Сорокина, В.И. Чирков (авторы адаптации). – Ярославль, НПЦ «Психодиагностика», 1996.

44. Франселла, Ф. Новый метод исследования личности / Ф. Франселла, Д. Баннистер. – М.: Прогресс, 1987.

45. Тестовая батарея общих профессиональных способностей (ТБОПС). Руководство // И.Г. Сенин, В.И. Чирков (авторы адаптации). – Ярославль, НПЦ «Психодиагностика», 1996.

46. Хьелл, Л. Теории личности. Основные положения, исследования и применение / пер. с англ. / Л. Хьелл, Д. Зиглер. – СПб, Питер Пресс, 1997.

47. Шмелев, А.Г. Психодиагностика личностных черт / А.Г. Шмелев. – СПб.: Речь, 2002.

48. Юнг, К.Г. Психологическая типология // Проблемы души нашего времени: пер. с нем. / К.Г. Юнг. – М.: Издат. группа «Прогресс», «Универс», 1994.

49. Ярошевский, М.Г. История психологии / М.Г. Ярошевский. – М.: Мысль, 1966.

50. Allport, G.W. Personality: A psychological interpretation / G.W. Allport. – N.Y.: Holt, Rinehart &Winston, 1937.

51. Cattell, R.B. Handbook of the Sixteen Personality factor Questionnair(16 PF) / R.B. Cattell, H.W. Eber, M.M. Tatsuoka. – Champaign: Institute for Personality and Ability testing, 1970.

52. Cattell, R.B. Advances in Cattellian personality theory. In L.A. Pervin (Ed.), Handbook of personality: theory and research / R.B. Cattell. – N.Y.: Guilford Press, 1990.

53. Cohen, R.J. Psychological testing. An introduction to tests and measurement / R.J. Cohen, P. Montague, L.S. Nathanson, M. Swerdlik. – Mayfield Publishing Company, California, 1988.

54. Costa, P.T. The NEO PI-R. Professional manual. Revised NEO Personality Inventory (NEO-PI-R) and Five factor Inventory (NEO-FFI) / P.T. Costa, Jr., & R.R. McCrae. – Odessa, Fl: Psychological Assessments Resources, 1992.

55. Costa, P. Personality development from adolescence through adulthood further cross– cultural comparisons of age differences

/ P. Costa, R. McCrae, T. Martin, V. Oryol, A. Rukavishnikov, I. Senin // In Temperament and Personality Development across the life span. Edited by V.J. Molfese, D.L. Molfese, LEA – New Jersey, London, 2000.

56. Cronbach, L.J. Essentials of Psychological Testing / L.J. Cronbach. – N.Y.: Harper & Row, 1984.

57. Domino, G. Psychological testing: An introduction / G. Domino. – Prentice-Hall, Inc, New Jersey, 2000.

58. Draguns, J. Personality Characteristics of the Nentsy in the Russian Arctic / J. Draguns, T. Martin, V. Oryol, A. Krilova, A. Rukavishnikov // American Behavioral Scientist. – 2000. – Vol. 44, № 1.

59. DuBois, P.H. A test-dominated society: China 1115 B.C.-1905 A.D. / P.H. DuBois // In A. Anastasy (Ed.) Testing problems in perspective. – Washington, D.C.: American Counsel on Education, 1966.

60. Eysenck, H.J. Biological dimensions of personality. In L.A. Pervin (Ed.), Handbook of personality: theory and research / H.J. Eysenck. – N.Y.: Guilford Press, 1990.

61. Kaplan, R.M. Psychological Testing: principal, applications, and issues / R.M. Kaplan, D.P. Saccuzzo. – Third Ed. Wadworth, Inc., Belmont, California, 1993.

62. Holland, J.L. Making vocational choices: A theory of vocational personality and work environments / J.L. Holland. – Englewood Cliffs, N.Y.: Prentic Hall, 1985.

63. Martin, T.A. (1997, August). Development of a Russian-language NEO-PI-R. Poster session presented at the 105th Annual Meeting of the American Psychological Association / T.A. Martin, J.G. Draguns, V.E. Oryol, I.G. Senin, A.A. Rukavishnikov, M.L. Klotz. – Chicago.

64. McCrae, R.R. Personality trait structure as a human universal / R.R. McCrae, & P.T. Costa, Jr. // America Psychologist. – 1997. – 52.

65. Meier, S.T. The chronic crisis in psychological measurement and assessment: a historical survey / S.T. Meier. – San Diego, N.Y., Boston, London and others; Academic Press., 1994.

66. Norman, W.T. Toward an adequate taxonomy of personality attributes / W.T. Norman // Journal of Abnormal and Social Psychology. –1963 – 66.

67. Roback, A.A. History of psychology and psychiatry / A.A. Roback. – N.Y., Philosofical Library, 1961.

68. Standarts for Educational and Psychological Testing. – American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1986.

69. Tiffin, J. Industrial Psychology / J. Tiffin, E. McCormick – Prentice-Hall, Inc., N.Y., 1965.

Приложения

Приложение 1

Критические значения коэффициента корреляции

Степень свободы (Объем выборки)	Уровень значимости		Степень свободы (Объем выборки)	Уровень значимости	
	5%	1%		5%	1%
20	0.42	0.54	45	0.29	0.37
21	0.41	0.53	50	0.27	0.35
22	0.40	0.52	60	0.25	0.33
23	0.40	0.51	70	0.23	0.30
24	0.39	0.50	80	0.22	0.28
25	0.38	0.49	90	0.21	0.27
26	0.37	0.48	100	0.20	0.25
27	0.37	0.47	125	0.17	0.23
28	0.36	0.46	150	0.16	0.21
29	0.36	0.46	200	0.14	0.18
30	0.35	0.45	300	0.11	0.15
35	0.33	0.42	400	0.10	0.13
40	0.30	0.39	500	0.09	0.12

Приложение 2

Критические значения t-критерия Стьюдента

Степень свободы (Объем выборки)	Уровень значимости		Степень свободы (Объем выборки)	Уровень значимости	
	5%	1%		5%	1%
10	2.23	3.17	27	2.05	2.77
11	2.20	3.11	28	2.05	2.76
12	2.18	3.05	29	2.05	2.76
13	2.16	3.01	30	2.04	2.75
14	2.14	2.98	40	2.02	2.70
15	2.13	2.95	60	2.00	2.65
16	2.12	2.92	80	1.99	2.64
17	2.11	2.90	100	1.98	2.63
24	2.06	2.80	120	1.98	2.62
25	2.06	2.79	200	1.97	2.60
26	2.06	2.78	500	1.97	2.59

Учебное издание

Орёл Валерий Емельянович
Сенин Иван Геннадьевич

Основы психодиагностики

Учебное пособие
Издание второе, переработанное и дополненное

Редактор, корректор А.А. Аладьева
Компьютерная верстка И.Н. Ивановой

Подписано в печать 02.08.2007. Формат 60x84/16. Бумага тип.
Усл. печ. л. 8,14. Уч.-изд. л. 6,1. Тираж 150 экз. Заказ

Оригинал-макет подготовлен
в редакционно-издательском отделе ЯрГУ.

Ярославский государственный университет
150000 Ярославль, ул. Советская, 14

Отпечатано
ООО «Ремдер» ЛР ИД № 06151 от 26.10.2001.
г. Ярославль, пр. Октября, 94, оф. 37
тел. (4852) 73-35-03, 58-03-48, факс 58-03-49.